

全国第六届研究生数学建模竞赛



题 目 基于向量自回归模型的城镇就业人数问题研究

摘 要：

本文在金融危机背景下，以宏观经济学理论为基础，结合实际情况和相关统计数据，建立多变量向量自回归模型研究我国城镇就业问题。

第一问，使用非参数 Spearman 秩和检验法和 Granger 因果检验对候选的 13 个因素与城镇就业人数进行检验，确定 GDP 总量、城镇总投资、教育投入、出口贸易额、城镇居民人均可支配收入等 10 个因素作为影响就业的主要因素，其中城镇就业人数与利率互为因果关系。然后对这 10 个因素进行路径分析，得到各变量间的层次和关联强度（见图 2）。

第二问，基于第一问中得到的重要因素，考虑城镇就业人数与因素间的关系，同时为了避免指标间多重共线性，本文建立了通用的多变量向量自回归对数线性模型，并给出了模型的求解方法。

第三问，选取合适指标，分别对包含主要的经济社会指标，浙江省，建筑业和农民工群体建立多变量向量自回归模型，并进行检验，发现所有模型的精度均达到 95% 以上。并且，通过多次对向量自回归模型的拟合，显示了该模型具有较强的稳定性和可操作性。

第四问，结合金融危机和四万亿投资背景，基于多变量向量自回归模型，得出三套仿真方案。首先得出只考虑金融危机影响的方案一。然后，考虑到国外投资对我国就业人数的特殊影响，建立了有效就业投资指标，以此指标作为基础得出仿真方案二。最后，引入金融危机影响函数 P 和投资影响函数 Z ，逐月对就业人数进行仿真。

第五问，首先，基于公平与效率的原则建立多目标规划模型，给出四万亿投资分配到各行业的最优方案。然后利用多变量向量自回归模型，从投资、出口、消费的角度探讨弥补就业缺口的对策，并提出增加第二、第三产业投资，刺激消费，扩大内需和增加投资等政策建议。

关键词：路径分析 多变量自回归 有效就业投资 就业缺口 多目标规划

参赛队号 1050401

队员姓名 龚梦 袁方 成峰

参赛密码 _____

(由组委会填写)

目 录

目 录	1
1 前 言	2
1.1 研究意义	2
1.2 国内外研究动态	2
1.3 问题的分析	3
1.4 基本假设	4
1.5 模型的符号说明	5
1.6 本文的架构设计	5
2 问题一解答——影响城镇就业的因素分析	6
2.1 指标的选取	6
2.2 数据来源及说明	7
2.3 对指标进行非参数 Spearman 秩和检验	9
2.4 对指标进行 Granger 因果检验	10
2.5 基于路径分析的指标相关强度分析	10
2.6 指标的最终确定及层次关系确定	13
3 问题二解答——建立城镇就业人数的数学模型	14
3.1 问题分析	14
3.2 经济指标间的信息反馈分析	14
3.3 多变量自回归对数线性模型的建立	14
3.4 确定多变量自回归对数线性模型的参数	15
3.5 模型系数求解算法	16
3.6 模型的检验与误差修正	16
4 问题三解答——建立多角度的城镇就业人数模型	18
4.1 建立包含经济社会指标的城镇就业人数模型	18
4.2 建立建筑行业城镇就业人数模型	21
4.3 建立浙江地区的城镇就业人数模型	25
4.4 建立农民工群体的城镇就业人数模型	29
5 问题四解答——仿真 2009 年及 2010 年上半年的就业人数	33
5.1 考虑经济危机的仿真方案	33
5.2 考虑国外投资特殊性的仿真方案二	34
5.3 考虑经济危机和投资变动的仿真方案三	37
6 问题五解答——提高城镇就业人口的建议	40
6.1 投资的分配	40
6.2 就业人数与投资、出口、和消费的分析	43
7 模型评价和推广	45
7.1 模型的优点	45
7.2 模型的缺点	45
7.3 模型的推广	45
参考文献	45
附录	47

1 前言

1.1 研究意义

在全球金融危机的背景下，我国的就业情况遭受着严峻的考验。众所周知，就业是重大的经济社会问题，也是重大的政治问题。它关系到人民群众的切身利益，关系到改革发展稳定的大局。如果解决不好，可能引发一系列社会矛盾。

随着金融危机的蔓延，对我国而言，解决当前大量的失业现状显得迫在眉睫。近年来，我国的经济增长一直处于较高水平，与此相反，就业状况则不容乐观，已经成为我国宏观经济发展过程中的一个比较突出的棘手问题。因此，对经济增长对就业影响做出系统的疏理，对于就业与经济增长的协调发展以及改善就业状况，进而提高社会福利水平都将是十分有意义的事情。

因此，在当前形势下，我们研究影响中国城镇就业的主要因素或指标，建立这些指标和城镇就业人数之间的数学模型，从包含主要的经济社会指标、分行业、分地区、分就业人群角度，建立更精确的数学模型来研究和预测各行业、各地区以及各就业人群的就业情况，并对2009年及2010年上半年的我国就业人数进行仿真，具有非常重要的现实意义。本文力求通过定量的研究，科学地给出提高我国城镇就业人数的有效建议，对指导我国严峻的就业问题具有非常重要的现实意义。

1.2 国内外研究动态

国外对于失业的研究涉及的方面很广。在研究劳动力市场机制对失业和就业的影响方面，Christopher Pissarides(1985)探究了劳动力市场模型，通过一个建立和破坏的结果，劳动力市场上的工人，和公司之间复杂的匹配和谈判过程，来解释劳动力市场上的失业。Dale Mortensen(1994)在文章中拓展了模型，使这个框架结构变得极有影响力。Horst Siebert(1997)认为70年代以来的制度变化是欧洲劳动力市场状况不佳的主要原因，并且对欧洲和美国的就业率差异进行了比较分析。Olivier Blanchard, Daniel Cohen和Cyril Nouveau(2005)分析了法国自从20世纪50年代开始的劳动力市场制度的演进和决定因素。在研究失业治理政策对失业影响方面，Cahuc和Kramarz(2004)提出了一套改善劳动力市场运行的改革方案，包括协调灵活性和就业稳定性的需求措施。Blanchard和Tirole(2003, 2005), Cahuc和Zylberberg(2005)建议成立一套结构性的失业税体系，使雇主将失业社会成本内部化。Brown等人(2006)研究了失业救济体系的长远改革以提高整个机制的效率。

我国学者对国外就业研究的成果也很显著。沈全水编写的著作《失业的出路——瑞典就业政策及其对中国的启示》主要对20世纪50年代到90年代初的瑞典就业政策进行了分析研究，强调国家调节与市场调节的结合，指出要推进中国市场导向的改革，我们可以从瑞典吸取哪些经验教训。由裘元伦、罗红波主编的《中国与欧洲联盟就业政策比较》中欧学者对劳动市场现状、影响劳动市场的各种因素、反失业及即或劳动市场的政策取向以及就业与福利保障制度进行了深入讨论，深受国内外读者的推崇。

国内学者对我国就业的研究也较多。围绕对经济增长与劳动力就业相关关系研究的文章各大报刊都有，这些研究为促进我国经济增长和劳动力就业提供了许多有益的措施和建议。目前，将理论和定性角度相结合的研究的成果为：李思（2007）通过对影响劳动力就业的因素进行分析提出实施以就业为中心的经济发展战略以及相应的措施；赵慧、刘德鑫、张家来（2007）从经济增长与劳动力就业的基本关系出发，在分析影响我国就业弹性深层次原因的基础上，提出了协调经济增长与扩大劳动力就业之间关系的政策建议；李俊锋、邹红美（2005）分析了我国经济增长和有效就业增长之间存在正向

的双向因果关系。然而,从省的层次来研究经济增长与劳动力就业相关关系的文献相对较少,王娟、熊建国(2007)分别对宁夏、安徽省经济增长与劳动力就业进行实证分析,从而提出了解决劳动力就业问题的建议。

在国内,使用定量的方法研究就业问题也越来越受到学者的青睐。张莉、张德然(2008)依据四川省近年来就业统计数据,通过建立多元线性回归模型和灰预测模型,从中选择优化模型进行分析和研究,力求对时下的就业有所借鉴。王婧(2008)首先利用因果分析理论从系统的角度对就业问题与众多宏观经济因素间的相互作用关系进行分析,进而通过改进的BP神经网络提出对就业总量进行宏观调控的理论和方法。肖春来(2009)拓展了CGE模型在劳动就业领域的新应用,建立了用于分析劳动就业的一般中国CGE模型,并且利用该模型从数量上模拟了产业结构调整 and 城乡劳动力转移带来的就业结构变化对中国宏观经济的影响,为实现劳动力资源的合理配置、实现经济增长与就业增长的良性互动提供政策参考依据。陈阳(2007)在讨论就业和经济增长之间关系的理论基础,分析了辽宁省就业弹性偏低和奥肯定律发生重大变异的机理,然后指出随时间变化的就业弹性系数才是分析辽宁省就业和经济增长之间关系的重要前提,最后构建辽宁省就业与经济增长之间关系的时变参数模型。

综观国内外现有的经济研究文献,经济学者们对劳动就业的研究在方法上明显缺乏交叉。就国外研究来看,大多数有关就业影响的文献,零星散存于经济学学者们的一些论文中,理论研究并不系统。从总体来看,我国学术界关于就业问题的研究尚显薄弱。国内学者在研究影响就业的因素和建立我国就业人数或者失业问题的数学模型时,选取的指标往往不够完善和全面,得到的模型没有通用性。因此,本文在力求克服以上缺陷的情况下,选取更完善的指标体系,建立能够通用的数学模型,来仿真和预测经济指标,并从定量的角度科学地提出提高我国城镇就业人口数的建议和意见。

1.3 问题的分析

就业是一个经济问题,更是一个社会问题。作为一个经济问题,就业是宏观经济中最重要的指标之一,是衡量企业和经济系统效率的重要标准;作为社会问题,就业是人生存和发展的基本前提,也是实现社会公平的基本条件。

1.3.1 宏观经济背景的分析

1.金融危机:由美国次贷危机引发的金融危机席卷全球,在全球经济一体化背景下,我国自然无法置身事外。金融危机对劳动力就业的冲击是最重要的问题之一。本文将重点研究金融危机对我国就业冲击和相应政策的研究。

2.四万亿投资:四万亿元的投资目的性很明确,就是要挖掘国内市场的潜力,扩大内需,保持经济的高增长,从而促进就业,解决由于金融危机造成的严重失业问题。本文将重点研究四万亿投资的效果和分配情况,并给出相应的政策建议。

1.3.2 数据的分析

由于本题没有提供足够数据,所以需要查询大量数据。我们主要数据来源于《中国统计年鉴》,《中国经济统计年鉴》,《中国劳动统计年鉴》,《中国海关统计年鉴》,《中国连锁经营年鉴》和政府官方网站等。本文在引用数据时均会明确标出数据出处。

1.3.3 失业的分析

根据宏观经济学理论和失业现象的直接原因,对失业进行分类,主要有摩擦性失业、结构性失业、周期性失业、季节性失业。结合金融危机的背景,农民工面对的主要是结构性和摩擦性的失业,而对于大学生而言,主要面对的是周期性失业和结构性失业。

1.3.4 投资的分析

为了便于研究和建模,本文将投资分为国外投资(FDI)和国内投资(DI)两类,并根据两

种投资不同方向和形式分别进行分析。本文在研究过程中，认为在短期，不考虑国内投资的乘数效应和挤出效应，不考虑国外投资的溢出效应。

1.3.5 技术进步的分析

相关研究表明(毕先萍, 赵坚毅《技术进步对我国就业总量及结构的影响》^[25]), 在短期, 技术进步会产生“创造性破坏”效应, 导致就业水平下降; 而在长期, 又会产生“补偿”效应, 产生新的就业岗位和工作机会。本文在仿真时间段中(2009年至2010年上半年), 不考虑此时间段的技术进步对就业的短期和长期的影响。

1.3.6 就业因素的分析

在现实生活中, 影响就业的主要因素或指标非常多。本文初步列举了5大指数和14个因素, 以便深入研究, 具体指数和因素如图:

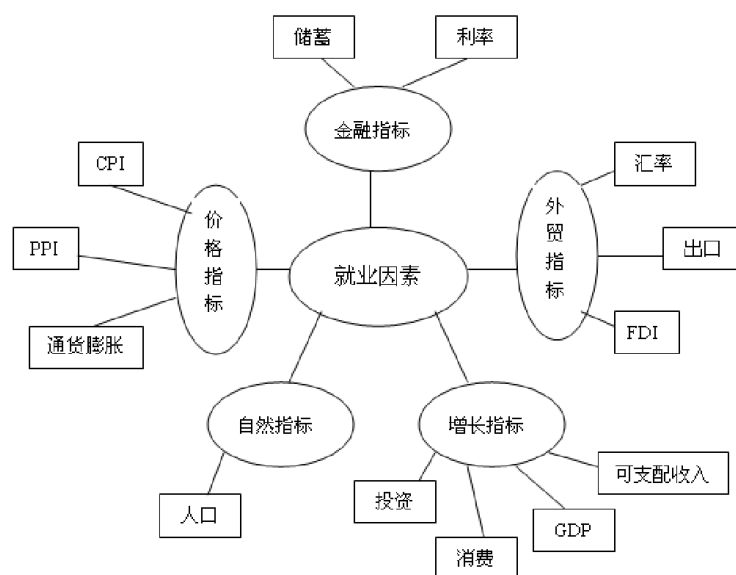


图1 指标体系关系图

1.3.7 隐性失业的分析

隐性失业就是具有劳动能力并在职工作但工作量不足, 不能通过工作获得社会认可的正常收入, 虽有工作岗位但未能充分发挥作用的失业, 或在自然经济环境里被掩盖的失业。许多发展经济学家认为, 我国失业的特点之一就是“隐性失业”。尤其金融危机的衰退时期, 由于企业开工不足, 即使未被解雇的工人也无法有效地使用。本文在使用就业人数统计数据的时候, 假设统计部门考虑了隐性失业人群。

1.4 基本假设

本文的基本假设如下:

- (1) 假设在 2009 年至 2010 上半年之间, 技术进步对就业不会有显著影响。
- (2) 假设在 2009 年至 2010 上半年之间, 无新的重大社会经济动荡和战争发生。
- (3) 假设在 2009 年至 2010 上半年之间, 我国各项宏观经济政策能及时准确地落实。
- (4) 假设在 2009 年至 2010 上半年之间, 我国对外经济政治政策无重大调整。
- (5) 假设在 2009 年至 2010 上半年之间, 我国对资本帐户管理政策无重大调整。
- (6) 假设在 2009 年至 2010 上半年之间, 无恶意国际炒家或机构扰乱金融市场。
- (7) 假设在 2009 年至 2010 上半年之间, 我国能对国际环境变化作出及时准确的反应。
- (8) 假设所有人为理性人。

1.5 模型的符号说明

符号	符号说明
y_i	第 i 年的城镇就业人数 单位：万人
x_{i1}	第 i 年的国内生产总值 (GDP) 单位：亿元
x_{i2}	第 i 年国家对于城镇的固定资产总投资 单位：亿元
x_{i3}	第 i 年国家对于居民的教育投入 单位：亿元
x_{i4}	第 i 年国家的出口贸易额 单位：亿元
x_{i5}	第 i 年城镇居民家庭人均可支配收入 单位：元
x_{i6}	第 i 年的通货膨胀率 单位：%
x_{i7}	第 i 年的消费指数 单位：以 1990 年为基年记为 100
x_{i8}	第 i 年的市场化程度 (以人均年社会消费品零售额表示) 单位：元/人
x_{i9}	第 i 年的城镇化程度 (以全国从事第三产业人口占全国人口比重表示) 单位：%
x_{i10}	第 i 年国家的总人口 单位：万人
x_{i11}	第 i 年国家的汇率 单位：%
x_{i12}	第 i 年国家的利率 单位：%
x_{i13}	第 i 年的居民储蓄总额 单位：亿元

1.6 本文的架构设计

文章的结构如下：

第一部分：前言，主要讲述了研究就业问题的研究意义，并对问题进行了重述。除此以外，还对国内外的研究现状进行简单的分析，对文章要注意的方面提出了基本假设。

第二部分：问题一解答。问题一主要选取了影响就业的主要因素或指标。该部分主要讲述指标的选取标准及原因，数据来源，非参数 Spearman 秩和检验方法运用及检验结果，对指标进行 Granger 因果检验的原理及结果，并用路径分析的方法，测出了各指标之间的相关强度。

第三部分：问题二解答。该部分主要是建立城镇就业人数与主要指标的数学模型。主要阐述了各经济指标间的信息反馈，并依据题意建立了差分方程，接着讲述了其参数确定及其检验。

第四部分：问题三解答。本文采用问题一中选取的重要指标，建立全国城镇就业人数的多变量自回归对数线性模型，然后分行业、分地区、分就业人群分别取一个样本进行分析求解。本文取建筑业、浙江省、农民工作为样本进行分析。

第五部分：问题四解答。在考虑不同情况下，建立三个模型，进行随机模拟。

第六部分：问题五解答。主要解决了两个重要问题：1、四万亿投资合理分配于不同行业，2、从投资、出口、消费的角度探讨弥补就业缺口的对策，并提出相依的建议。

2 问题一解答

——影响城镇就业的因素分析

2.1 指标的选取

根据相关研究和红股经济学理论,选取了 13 个可能影响就业的指标。它们分别是 GDP 总量、城镇总投资、教育投入(以教育投入资金表示)、出口贸易额、城镇居民人均可支配收入、通货膨胀率、消费指数、市场化程度(以人均年社会消费品零售额表示)、城镇化水平(以全国从事第二三产业人口占全国人口比重表示)、全国人口总数、居民储蓄总额、汇率以及利率。

1.GDP 总量与就业的关系。

GDP 即国内生产总值,从产品形态上看,表现为一国所有常住单位在一定时期内提供的最终产品的价值之和。这里的最终产品是指直接用于生产消费和生活消费的投资品和消费品。增加 GDP 有助于改善民生。而就业是民生之本。实践证明,GDP 增长也有助于带动就业的增加,就业与 GDP 之间存在相关关系,尽管这种关系在不同时期是不同的。“九五”时期平均就业弹性系数为 0.14,“十五”时期为 0.12,2008 年为 0.08,即 GDP 每增长一个百分点仅能带动 80 万个就业。这表明,我国经济增长与就业增长存在相关性,增长能够带来就业机会的增加(陈享光,2009)。

2.城镇总投资与就业的关系。

根据经济学理论,我们知道政府对城镇的总投资越多,就会有更多的项目带动就业,劳动密集型产业特别是中小企业和服务业的发展,会吸纳较多的就业。因此,城镇总投资和就业之间应该有正影响。

3.教育投入与就业的关系。

在本文中,教育投入是以教育投入资金来衡量的。国家对教育的投入越大,就业人员的素质就越高,接受教育的机会越多,就业前景越广阔。因此,在一定程度上来说,国家对教育的投入影响了我国的就业。

4.出口贸易额与就业的关系。

近年来我国贸易持续顺差,进口额和出口额增加幅度都比较大,对外依存度有所提高,对国民经济拉动发挥了重要作用。我国沿海地区依靠廉价的劳动力和资金、技术有效结合,大力发展外向型经济,极大地吸收了劳动力就业。就现状分析而言,出口贸易额与就业正相关,即出口额越大,我国就业人数就越多。

5.城镇居民人均可支配收入与就业的关系。

大力开发就业岗位,城镇新增就业人数就会伴随着增加,就业规模的扩大将极大地推动城镇劳动力就业,同时也会提高城镇居民的人均可支配收入,人民生活将改善。故城镇居民人均可支配收入与就业存在着联系。

6.通货膨胀率与就业的关系。

根据菲利普斯曲线,通货膨胀与失业率是成替代关系的.菲利普斯曲线本义是关于货币工资变动率和失业率之间关系的一条曲线.1960 年萨缪尔森与索洛在《反通货膨胀政策的分析》一文中用通货膨胀率替换了货币工资增长率。从而将原始菲利普斯曲线描述的关系延伸为失业率与通货膨胀率之间的反向关系:失业率高时.通货膨胀率就低,通货膨胀率高时.失业率就低.而失业率和就业人数有关,所以认为通货膨胀率与就业存在着联系。

7.消费指数(CPI)与就业的关系。

根据宏观经济学理论和各国经济变化的实践,物价水平和失业率存在着此消彼长的关系,即所谓菲利普曲线,它描述了通货膨胀和失业率之间的替代关系;物价涨则就业率升,物价低则就业率低。可见,消费指数 CPI 对就业人数有影响。

8.市场化程度与就业的关系。

市场化程度以人均年社会消费品零售额表示。市场化程度越高即人均年社会消费品零售额就越高。人均年社会消费品零售额反应了居民的人均消费水平,居民的消费水平又和就业紧紧联系。所以,市场化程度对就业有影响。

9.城镇化水平与就业的关系。

城镇化水平以全国从事第二三产业人口占全国人口的比重来衡量。城镇化水平的提高,可以带动第三产业特别是服务业的发展,而服务业是吸收城镇劳动力就业的一大行业。故城镇化水平对就业有影响。

10.人口总数与就业的关系。

从直观上看,人口越多,国家面临的就业问题就越大。人口总数不同,各行业,各地区,各就业人群所承受的就业压力就不一样。一般来说,人口越少,就业压力越小;人口越多,就业压力越大。故人口总数对就业有影响。

11.居民储蓄总额与就业的关系。

我国的经济增长率是高增长高储蓄。我国从改革开放至今,经济的平均增长率接近10%,能够保持如此高的经济增长速度,靠的是世界上最高的储蓄率来支撑。这是我国经济有强大发展动力的主要原因,同时也为经济进一步发展提供了充足的资金来源,为投资及就业提供了更多的机会。因此,居民储蓄总额对就业有影响。

12.汇率与就业的关系。

近年来,随着国际贸易保护主义重新兴起,国际金融环境发生改变,对中国来说,表现为人民币被迫升值。理论上,一国汇率升值会使本国产品出口减少,从而导致本国出口部门吸收就业的能力减弱。1994-2004 年的汇率与就业数据显示,人民币升值与我国就业之间呈现出很强的正相关关系。此外,我国出口贸易部门就业弹性远小于非出口贸易部门的就业弹性,汇率升值对非出口贸易部门的就业增加率远大于出口贸易部门的失业率。因此,人民币升值对整个就业,可能有有利的一面(孙恩杰,2008)。故汇率对就业有影响。

13.利率与就业的关系。

如果利率升高,人们就倾向于储蓄,而投资将减少。由于要承担更多的利息,企业也不愿意去贷款扩大经营规模,企业吸纳的城镇劳动力就业人数减少,失业率就会提高。故利率对就业有影响。

2.2 数据来源及说明

本文选取了 1990-2006 年影响我国就业的 13 个指标的时间序列数据,见表 1 以及表 1(续)。

表 1 城镇就业人数与各影响因素数据一览表

年份	城镇就业人数	GDP 总量	城镇总投资	教育投入	出口贸易额	城镇可支配收入	通货膨胀率
	万人	亿元	亿元	亿元	亿元	元	%
1990	17041	18547.9	6767.2	102.52	1510.2	1510.2	3.1
1991	17465	20250.4	8542.5	119.39	1700.6	1700.6	3.4
1992	17861	23134.2	10317.8	151.34	2026.6	2026.6	6.4
1993	18262	26364.7	12093.1	204.45	2577.4	2577.4	14.7
1994	18653	29813.4	13868.4	261.53	3496.2	3496.2	24.1

1995	19040	33070.5	15643.7	352.69	4283	4283	17.1
1996	19922	36380.4	17567.2	430.29	4838.9	4838.9	8.3
1997	20781	39762.7	19194.2	529.21	5160.3	5160.3	2.8
1998	21616	42877.4	22491.4	615.56	5425.1	5425.1	-0.8
1999	22412	46144.6	23732.0	711.65	5854.02	5854.02	-1.4
2000	23151	50035.2	26221.8	823.37	6280	6280	0.4
2001	23940	107449.7	30001.2	951.36	6859.6	6859.6	0.7
2002	24780	117208.3	35488.8	1164.56	7702.8	7702.8	-0.8
2003	25639	128958.9	45811.7	1671.14	8472.2	8472.2	1.2
2004	26476	141964.5	59028.2	2024.82	9421.6	9421.6	3.9
2005	27331	156775.3	75095.1	2209.199	10493	10493	1.8
2006	28310	204556.1	93368.7	2270.237	11759.5	11759.5	1.5
2007	29350	26651.01	117464.5	2375.557	13785.8	13785.8	4.8

其中, 指标 GDP 总量、城镇总投资、教育投入(以教育投入资金表示)、出口贸易额、城镇居民人均可支配收入、通货膨胀率、消费指数、市场化程度(以人均年社会消费品零售额表示)、城镇化水平(以全国从事第二三产业人口占全国人口比重表示)、全国人口总数和居民储蓄总额的数据来源于 1991-2007 年的《中国统计年鉴》, 而指标利率和汇率来源于《世界经济年鉴 2008-2009》。城镇就业人数与各影响因素的数据如表 1 所示。

表 1 (续) 城镇就业人数与各影响因素数据一览表

年 份	消费指数 1990 为 100	市场化程度 元/人	城镇化水平 %	人口总数 万人	汇率 万人	利率 %	居民储蓄总额 亿元
1990	103.1	699.75	22.01	114333	478.32	1.8	5911.2
1991	103.4	812.96	22.79	115823	478.32	1.8	9241.6
1992	106.4	938.29	23.43	117171	478.32	3.15	11759.4
1993	114.7	1051.5	24.58	118517	478.32	3.15	15203.5
1994	124.1	1357.1	25.72	119850	478.32	3.15	21518.8
1995	117.1	1702.4	26.86	121121	835.1	1.98	29662.3
1996	108.3	2024.2	27.89	122389	831.42	1.71	38520.8
1997	102.8	2208.2	28.29	123626	828.98	1.44	46279.8
1998	99.2	2336.7	28.42	124761	827.91	1.44	53407.5
1999	98.6	2475.2	28.32	125786	827.83	0.99	59621.8
2000	100.4	2694.7	28.44	126743	827.84	0.99	64332.4
2001	100.7	2945.7	28.61	127627	827.7	0.99	73762.4
2002	99.2	3184.9	28.72	128453	827.7	0.72	86910.6
2003	101.17	3558	29.32	129227	827.7	0.72	103617.3
2004	103.9	4163	30.72	129988	827.68	0.72	119555.4
2005	101.8	5153	31.96	130756	819.17	0.72	141051
2006	101.45	5828	33.35	131448	797.18	0.72	161587.3
2007	104.8	6769	34.47	132129	760.4	0.72	172534.2

资料来源: 利率和汇率来源于《世界经济年鉴 2008-2009》, 其它指标来源于 1991-2008 年的《中国统计年鉴》。

2.3 对指标进行非参数 Spearman 秩和检验

斯皮尔曼（Spearman）相关系数是描述两组变量之间是否存在相同或相反趋同性的一种指标，由于该检验不需要假定服从正态分布，仅需要确定变量在每个点（时期）上的等级即可获得，因此具有较好的性质。在两组数据都没有重复观测值的情况下，斯皮尔曼等级相关系数的公式为：

$$r_{SP} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

其中 d_i 表示两组数据的等级之差， n 为样本量。

2.3.1 斯皮尔曼等级相关系数的检验

和其它的推断一样，当以样本的数据来推测总体时，由于样本带有随机性，在小样本时数据间有相关，但总体之间不一定相关。因此有必要进行假设检验。在此，设定原假设 H_0 ：研究的总体之间无相关（即 $\rho = 0$ ），备责假设为 H_1 ：研究的总体之间有相关（即 $\rho \neq 0$ ）。

检验的样本估计量为样本的相关系数 r_{SP} ，在小样本的情况下通常临界值的 r 可直接查表，但在大样本的情况下可以通过变换

$$t = r_{SP} \sqrt{\frac{n-2}{1-r_{SP}^2}}$$

服从 $t(n-2)$ 的 t 分布，采用 t 检验。运用 SAS 软件，得到城镇就业人数与各指标间的斯皮尔曼检验结果如表 2

表 2 斯皮尔曼检验结果

变量	$\sum d_i^2$	r_{SP}	t 检验 P 值	趋同性说明
城镇就业人数与 GDP 总量	182	0.81218	<.0001	正向趋同
城镇就业人数与城镇总投资	0	1.00000	<.0001	完全正向趋同
城镇就业人数与教育投入	0	1.00000	<.0001	完全正向趋同
城镇就业人数与出口贸易额	0	1.00000	<.0001	完全正向趋同
城镇就业人数与城镇居民人均可支配收入	0	1.00000	<.0001	完全正向趋同
城镇就业人数与通货膨胀率	1376	-0.42024	0.0825	没有关联
城镇就业人数与消费指数	1376	-0.42024	0.0825	没有关联
城镇就业人数与市场化程度	0	1.00000	<.0001	完全正向趋同
城镇就业人数与城镇化水平	2	0.99794	<.0001	正向趋同
城镇就业人数与人口总数	0	1.00000	<.0001	完全正向趋同
城镇就业人数与汇率	739	0.23722	0.3432	没有关联
城镇就业人数与利率	1866	-0.92578	<.0001	负向趋同
城镇就业人数与居民储蓄	0	1.00000	<.0001	完全正向趋同

由以上斯皮尔曼检验的结果可以看出，城镇就业人数与通货膨胀率、消费指数以及汇率没有关联。而城镇就业人数与城镇总投资、劳动力素质、出口贸易额、城镇居民人均可支配收入、市场化程度、人口总数以及居民储蓄具有完全的正向趋同，关系非常紧密。对于指标 GDP 总量和城镇化水平，与城镇就业人数呈正向趋同，仅有利率与城镇就业人数呈负向趋同。

通过斯皮尔曼检验,我们剔除掉通货膨胀率、消费指数以及汇率三个无关联的指标。剩下的 10 个指标分别是 GDP 总量、城镇总投资、教育投入、出口贸易额、城镇居民人均可支配收入、市场化程度、城镇化水平、人口总数、利率以及居民储蓄。这 10 个指标将作为影响就业的主要因素或指标。

2.4 对指标进行 Granger 因果检验

接着,对这 10 个指标与城镇就业人数之间的因果关系进行辨别与检验。为此,本文运用 Granger (格兰杰) 因果关系检验法对各指标及城镇就业人数之间进行检验,结果表明各数据皆为平稳性数据。

在给定显著性水平 $\alpha=5\%$ 的情况下, GDP 总量是城镇就业人数的原因; 城镇总投资是城镇就业人数的原因; 教育投入是城镇就业人数的原因; 出口贸易额是城镇就业人数的原因; 城镇居民可支配收入是城镇就业人数的原因; 市场化程度是城镇就业人数的原因; 城镇化水平是城镇就业人数的原因; 人口数是城镇就业人数的原因; 居民储蓄是城镇就业人数的原因。其中值得注意的是, 城镇就业人数与利率互为因果, 即利率是城镇就业人数的原因, 且城镇就业人数是利率的原因。

Granger 检验结果如表 3 所示:

表 3 Granger 因果检验结果

因果检验原假设	Obs	F-Statistic	Probability	结果
城镇就业人数不是 GDP 总量的原因	16	1.25044	0.28229	GDP 总量是因
GDP 总量不是城镇就业人数的原因		5.14857	0.03960	
城镇就业人数不是城镇总投资的原因	15	0.32710	0.72844	城镇总投资是因
城镇总投资不是城镇就业人数的原因		2.76638	0.09016	
城镇就业人数不是教育投入的原因	15	0.06810	0.93456	教育投入是因
教育投入不是城镇就业人数的原因		4.15882	0.04518	
城镇就业人数不是出口贸易额的原因	15	0.38271	0.54684	出口贸易额是因
出口贸易额不是城镇就业人数的原因		5.90847	0.03029	
城镇就业人数不是城镇居民可支配收入的原因	15	0.74576	0.55444	城镇居民可支配收入是因
城镇居民可支配收入不是城镇就业人数的原因		4.66144	0.03630	
城镇就业人数不是市场化程度的原因	15	1.53074	0.25911	市场化程度是因
市场化程度不是城镇就业人数的原因		37.8583	1.2E-05	
城镇就业人数不是城镇化水平的原因	16	0.59357	0.63656	城镇化水平是因
城镇化水平不是城镇就业人数的原因		1.50462	0.08579	
城镇就业人数不是人口总数的原因	16	4.34593	0.04065	人口数是因
不是城镇就业人数的原因		7.45144	0.00900	
城镇就业人数不是利率的原因	16	4.31963	0.04125	互为因果
利率不是城镇就业人数的原因		4.97629	0.02890	
城镇就业人数不是居民储蓄的原因	16	0.61474	0.55836	居民储蓄是因
居民储蓄不是城镇就业人数的原因		8.86873	0.00508	

为了考查各指标相对于城镇就业人数的重要程度, 本文将接着对通过斯皮尔曼检验的指标进行路径分析。

2.5 基于路径分析的指标相关强度分析

通过使用 Spearman 相关系数检验, 得到经过筛选的影响就业人数的 10 个指标。但

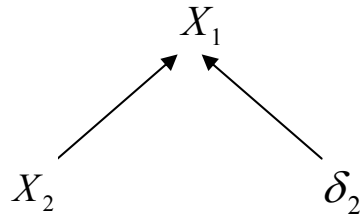
是这些指标之间存在着潜在的相互关系和层次关系，所以这里我们使用路径分析方法来确定各指标之间的多层因果关系及其相关强度。

2.5.1 路径分析模型的介绍

路径分析模型是由美国遗传学家 Swell Wright 于 1921 年首创，用于研究多个变量之间多层因果关系及其相关强度的方法，现已发展成为社会学的主要研究方法之一。路径分析的主要目的是检验一个假想的因果模型的准确和可靠程度，测量变量间因果关系的强弱，回答下述问题：①模型中两变量 x_j 与 x_i 间是否存在相关关系；②若存在相关关系，则进一步研究两者间是否有因果关系；③若 x_j 影响 x_i ，那么 x_j 是直接影响 x_i ，还是通过中介变量间接影响或两种情况都有；④直接影响与间接影响两者大小如何。

该模型实际上就是一特定的网络图形描述的标量之间的线性关系的假设，模型的建立必须经过理论论证，将总的效应分解成为直接效应、间接效应和联合效应，然后将有关变量按照推理的层次和因果关系顺序排列，然后按照结果模型画出路径图。

当一个变量 X_1 在时间上先于另一变量 X_2 时，可以假设 X_1 引起 X_2 ，若用图形表示可以写为 $X_1 \rightarrow X_2$ ，在这种关系中允许存在误差 δ_2 ，于是其路径图是：



借助一个线性模型

$$X_2 = \beta_0 + \beta_1 X_1 + \delta_2$$

可用标准化的的使用和常用记号写为

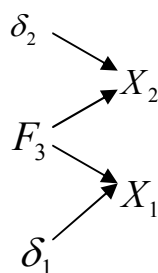
$$\frac{X_2 - \mu_2}{\sqrt{\sigma_{22}}} = \beta_1 \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \left(\frac{X_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) + \sqrt{\frac{\sigma_{\delta\delta}}{\sigma_{22}}} \frac{\delta_2}{\sqrt{\sigma_{22}}}$$

这里就连标准误差 δ 也有一个系数，通常把标准化模型中的参数 p 称为路径系数。上式中的因果关系模型意味着

$$\begin{aligned} \rho_{12} &= \text{Corr}(X_2, X_1) = \text{Corr}(Z_2, Z_1) = p_{21} \\ \text{Var}(Z_2) &= p_{21}^2 \text{Var}(Z_1) + p_{2\delta}^2 \text{Var}(\delta) = p_{21}^2 + p_{2\delta}^2 = 1 \end{aligned}$$

第二个方程表明说所假设的路径的图是直包含的，或由所示的变量可完全决定，因为 Z 的方差和为 1。

从数学上说，假定 X_2 引起 X_1 ，或者假定存在一个新的公共因子（例如 F_3 ）以解释 X_1 和 X_2 之间的相关性，这同样是合乎逻辑的，在后一种情况中， X_1 与 X_2 之间是伪相关的（不是因果相关）。其路径是：



这里我们允许在这个关系中再次出现误差，利用标准化变量，上面路径图的线性模型成为

$$Z_1 = p_{13}F_3 + p_{1\delta_1}\delta_1$$

$$Z_2 = p_{23}F_3 + p_{2\delta_2}\delta_2$$

其中标准化误差 δ_1 和 δ_2 互不相关，与其他变量以及 F_3 也不相关，因此，路径系数和相关系数有以下关系：

$$\rho_{12} = \text{Corr}(X_1, X_2) = \text{Corr}(Z_1, Z_2) = p_{13}p_{23}$$

$$\text{Var}(Z_1) = p_{13}^2 + p_{1\delta_1}^2 = 1$$

$$\text{Var}(Z_2) = p_{23}^2 + p_{2\delta_2}^2 = 1$$

和

$$\rho_{13} = \text{Corr}(Z_1, F_3) = p_{13}$$

$$\rho_{23} = \text{Corr}(Z_2, F_3) = p_{23}$$

路径分析包括两个主要部分：（1）路径图；（2）分解相关系数为简单路径和负荷路径的路径系数之和。这些特征使我们能够直接或简洁地去度量一个变量对另一个变量的效应。

2.5.2 路径分析模型的建立与求解

现在以全国统计数据为例，以城镇就业人数与各影响因素的相关系数矩阵为基础，建立路径分析模型，利用 SAS 计算得到如下路径图：

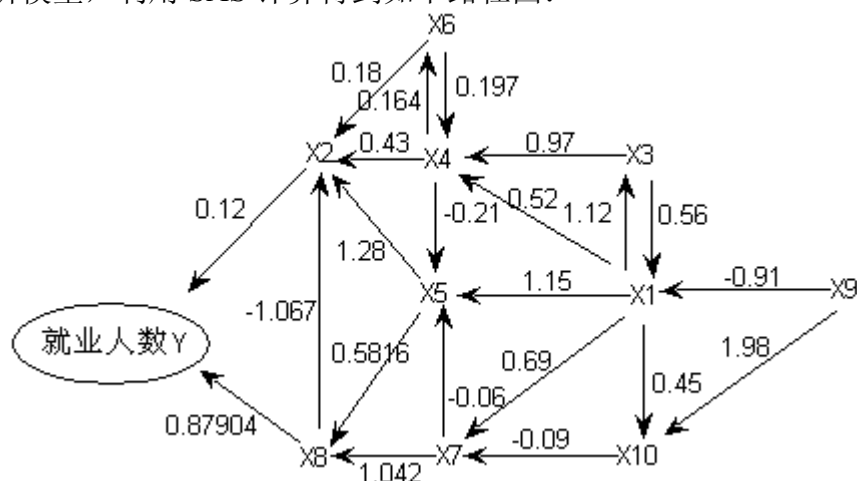


图 2 城镇就业人数与各影响因素的相关路径图

Y ——城镇就业人数

X_2 ——城镇固定资产投资

X_1 ——全国 GDP 总量

X_3 ——教育投入

X_4 ——年出口总额
 X_5 ——城镇居民可支配收入
 X_6 ——城镇市场化水平
 X_7 ——城镇化水平
 X_8 ——全国人口总数
 X_9 ——利率
 X_{10} ——城镇居民储蓄总额

由于篇幅有限，本文在此处列举三个模式的计算结果。

(1)模式一（城镇就业人数路径）计算结果：

Variable	ParameterEstimate	Standard Error	Type II SS	F Value	Pr> F
X2	0.12448	0.01681	0.01631	54.83	<.0001
X8	0.87904	0.01681	0.81328	2734.32	<.0001

(2)模式二（城镇固定资产投资）计算结果：

Variable	ParameterEstimate	Standard Error	Type II SS	F Value	Pr> F
X4	0.4317	0.11317	0.00605	13.03	0.0041
X5	1.2804	0.11760	0.00530	11.41	0.0062
X6	0.1832	0.04992	0.00309	6.66	0.0255

(3)模式三（年出口总额）计算结果：

Variable	ParameterEstimate	Standard Error	Type II SS	F Value	Pr> F
X1	0.5214	0.01681	0.01631	54.83	<.0001
X3	0.9731	0.01681	0.81328	2734.32	<.0001

2.5.3 路径分析模型的结果分析

通过路径图可以清晰明了的看出各因素的层次和相互作用的关系。城镇固定资产投资和全国人口总数，对城镇就业人数有显著的影响，增加投资和控制人口数目可以有效的提高城镇就业人数。以此类推，对于城镇固定资产投资来说，年出口总额、利率、城镇居民可支配收入和城镇市场化水平均对其有显著影响。

2.6 指标的最终确定及层次关系确定

本节中通过查阅资料，我们获取了 1990-2007 年影响就业的 13 个指标的基本统计数据，并对数据进行了处理。通过问题的分析及合理的假设，先运用非参数 Spearman 秩和检验法分别对我国城镇就业人口数与各指标之间的相关性进行检验，将不能通过检验的指标剔除，10 个指标通过非参数 Spearman 秩和检验。

事实证明，能通过数学上的非参数 Spearman 秩和检验，并在经济理论上合理的影响就业的主要指标分别是 GDP 总量、城镇总投资、教育投入、出口贸易额、城镇居民家庭人均可支配收入、市场化程度、城镇化水平、人口总数、利率以及居民储蓄。

另外，本文利用路径分析对各指标对城镇就业人口数的相关程度进行了分析。分析的结果为：城镇固定资产投资和全国人口总数，对城镇就业人数有显著的影响，增加投资和控制人口数目可以有效的提高城镇就业人数。对于城镇固定资产投资来说，年出口总额、利率、城镇居民可支配收入和城镇市场化水平均对其有显著影响。

3 问题二解答-

——建立城镇就业人数的数学模型

3.1 问题分析

本节要建立城镇就业人数与问题一中选取的主要因素或指标之间联系的数学模型。对于不同的实际问题可能需要不同的指标体系，要想将所研究的问题推广，使其具有更高的通用性，需要建立一个通用的数学模型。该模型中的指标数目可增可减，可以适用于多种实际情况。本文将建立城镇就业人数与 n 个就业指标间的数学模型。

3.2 经济指标间的信息反馈分析

在建立就业人数与主要经济指标的联系时，不可避免地要考虑指标间的相互作用。指标间的相互作用，往往会干扰我们所研究的问题。所以，本文也将分析指标之间的关系作为一项重要的内容。

本文结合实际情况，应用信号系统的反馈系统知识与微分-差分思想，综合考虑指标间的相互作用。为了更清晰地反应文中考虑指标间相互关系的作用，以下给出指标间的作用如图 2：

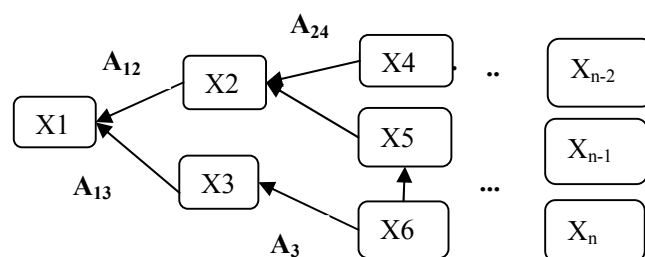


图 3 指标关系图

该图可以明显地显示指标间深层的内部联系，对我们解决问题，分析问题，建立方程具有重要的作用。

3.3 多变量自回归对数线性模型的建立

根据上节所述，在实际问题中，不同指标之间一般是互相影响的。本节在考虑指标间互相影响的情况下，建立了以时间递推的差分方程。

建立方程中的指标为： x_1, x_2, \dots, x_n 。指标在时间 t 时刻的数值为 $x_i(t)$ 。由此，根据指标间互相影响认为 $t+1$ 时刻的数值由 t 时刻的各个指标所决定，但是考虑到消除变量间的多重共线性，本文先对原数据取对数，然后运用处理后的数据建立线性对数模型，对于单个变量可以得到如下方程：

$$\ln x_k(t+1) = m_{k1} \ln x_1(t) + \dots m_{kn} \ln x_n(t) + \varepsilon$$

其中， $k=1..n$ 。由上方程于是可以得到 n 个表示所有内生变量 $x_i(t)$ 以及外生变量 $y_i(t)$ 指标间变化规律的方程，综合所有方程可以到的到一个方程组如下：

该方程组即为本文通用的模型, 但该模型的建立依赖于差分方程中合理的系数确定, 保证模型没有显著误差, 本文采取最小二乘法求得系数的近似精确解, 并对模型的可信性进行论证。

通过前面的步骤,可以得到综合的差分方程组。通过各种方法的比较,本文选取最小二乘法来确定方程组的系数。本文利用解超定方程组来获得其最小二乘解。

设线性方程组 $Ax=b$ 中，A 矩阵是 m 年 n 个指标的实际历史数据。b 为应变量的历史数据，x 为方程的解。当 $m>n$ 时，方程组中方程个数多于未知数个数，此时就称为超定方程组，一般来说超定方程组是无解的，为此我们要求一个近似解因此用最小二乘法。

证明：先证必要性

多元函数 f 求极限的必要条件如下:

再将以上方程整理得到矩阵分解式:

然后转化为矩阵形式为： $A'Ax = A'b$ 。

已知 x 为 $A'Ax = A'b$ 的解, 要证 x 为最小二乘解。

任取向量 $\bar{x} \neq x$, $y = \bar{x} - x$, 且 $y \neq 0$ 。于是有以下推导:

所以, x 是方程的最小解。

3.4.3 超定方程组的构造

在已知一段时间的历史数据的情况下，根据已经建立的差分方程模型，可以求系数的超定方程组。

其中，方程常数项 $b = (x_k(2), x_k(3), \dots, x_k(t+1), \dots, x_k(m))'$ 表示被影响的指标数值。对应着方程的系数矩阵 A 为：

$$\begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(m-1) & x_2(m-1) & \cdots & x_n(m-1) \end{bmatrix}$$

于是，可以根据以上得到的数据，利用最小二乘法就可以得到模型中所有系数，这样便可以得到完整的模型。

3.5 模型系数求解算法

由上文的分析论证，可以得到计算差分方程组系数的方法。由此，下文给出计算差分方程系数的计算公式和步骤：

对于每个方程的求解有相同的方法，每一次计算一个差分方程的系数，步骤如下：

Step1: 根据搜索得到的数据，可以得到不同时刻不同指标的数值矩阵 A 、 b 。

Step2: 根据第一步得到的数据，建立超定方程组，使用 MATLAB 软件求解差分方程组的系数，计算系数公式： $m = (A'A)^{-1} \times Ab$ 。

Step3: 记录计算结果， $i=i+1$ ，进入下一个差分方程组的计算。判断 $i < n$ ，若是则转入 Step1，否则转入 Step4。

Step4: 输出结果，作图比较拟合效果。

计算差分方程组算法如流程图 3：

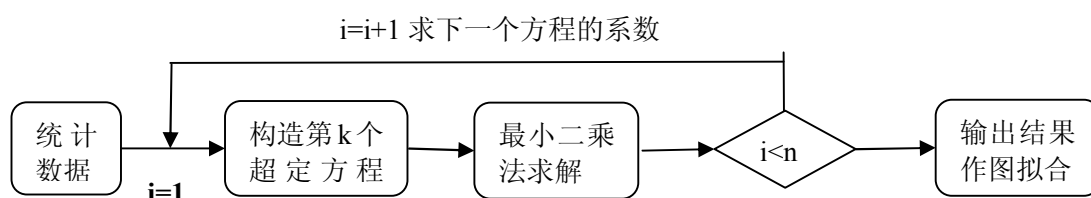


图 4 系数计算流程图

通过以上算法，编写 MATLAB 程序就可以实现差分方程系数的求解。

3.6 模型的检验与误差修正

由于多变量的 VAR 模型是一种非理论性的模型，它无需对变量作任何先验性约束。因此，在分析该模型时本文采用了脉冲响应函数、方差分解两种方法对模型进行检验。又因为非稳定的 VAR 模型不能做脉冲响应函数分析，所以首先要做平稳性分析。最后，还要对模型的拟合做残差检验，本文使用 Johansen 协整检验。

3.6.1 平稳性检验

可以利用 EViews 软件，调用 AR Boots Table 和 AR Boots Graph 模块，可以得到根表和根图。观察根表如果全部根的倒数都在单位圆之内，则说明 VAR 模型是稳定的。

3.6.2 脉冲响应函数

脉冲响应函数是分析误差项(扰动项)发生变化或者模型受到某种冲击时对系统的动态影响。对于方程组中的扰动项 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)'$ ，假定是具有下面这样性质的白噪声向量：

$$1、E(\varepsilon_i) = 0 \quad \text{对于 } \forall i, \quad i = 1, 2, \dots, n$$

$$2、\text{var}(\varepsilon) = E(\varepsilon' \varepsilon) = \Sigma.$$

对于本题多变量的 VAR 模型，可以得到 $VMA(\infty)$ 表达式(其中 L 为滞后算子,参考高铁梅《计量经济分析方法与建模》附件^[5] p250-278):

$$y_i = \sum_{j=1}^n (c_{ij}^{(0)} \varepsilon_j + c_{ij}^{(1)} \varepsilon_j + \dots + c_{ij}^{(n)} \varepsilon_j) \quad i=1..m$$

使用 EViews 软件作出多图表脉冲响应函数，观察函数因内生变量在序列冲击时随着时间的推移影响是否越来越小，若趋于零则说明 VAR 模型是一个稳定的系统。否则，认为不稳定。

3.6.3 方差分解

方差分解是通过分析每一个结构冲击对内生变量的变化(用方差来度量)的贡献率，进一步的评价不同结构冲击的重要性。因此，方差分解给出对 VAR 模型中的变量产生影响的每一个随机扰动的相对重要性的信息。

在 $VMA(\infty)$ 基础上提出方差分解，定量的但是相当粗糙的把握变量间的影响关系，由 Sims 于 1980 年提出，给出近似的相对方差共献率(RVC):

$$RVC_{j \rightarrow i}(s) = \frac{\sum_{q=0}^{s-1} (c_{ij}^q)^2 \sigma_{jj}}{\sum_{j=1}^k \sum_{q=0}^{s-1} (c_{ij}^q)^2 \sigma_{jj}}, \quad i, j = 1, 2, \dots, k$$

其中 $RVC_{j \rightarrow i}(s)$ 大时，意味着第 j 个变量对第 i 个变量的影响大，相反较小时，可以认为第 j 个变量对第 i 个变量的影响小。

3.6.4 Johansen 协整检验

协整检验是一种对模型残差的检验，其中本文使用的 Johansen 协整检验是基于回归系数的协整检验。

Johansen 协整检验中要做有关协积要做协积得假设检验，然后使用 EViews 软件分析，检验的结果是分层输出，不同行的假设是否通过检验可以知道不同变量之间是否存在均衡关系。

4 问题三解答

——建立多角度的城镇就业人数模型

4.1 建立包含经济社会指标的城镇就业人数模型

本文力求建立包含经济社会指标的城镇就业人数模型。研究包含经济社会指标的城镇就业人数，能够从经济和社会角度来分析对城镇就业人数的影响。

4.1.1 影响中国城镇就业的指标选取

影响全国城镇就业的主要因素或指标很多。模型中的因变量为全国城镇就业人数，影响就业的指标则选取了 8 个主要指标，分别是 GDP 总值、国家总投资、教育经费、出口贸易额、市场化程度、人口总量、利率以及城镇化程度。通过查阅数据，得到了所有指标从 1990-2006 年的时间序列数据。如表 4 所示。

表 4 全国城镇就业人数与各影响因素数据一览表

年份	城镇就业人数	GDP 总量	国家总投资	教育经费	出口贸易额	市场化程度	人口总量	利率	城镇化程度
	万人	亿元	亿元	亿元	亿元	元/人	万人	%	%
1990	17041	18547.9	4517	102.52	1510.2	-113.21	114333	1.8	22.01
1991	17465	20250.4	5594.5	119.39	1700.6	812.96	115823	1.8	22.79
1992	17861	23134.2	8080.1	151.34	2026.6	938.29	117171	3.15	23.43
1993	18262	26364.7	13072.3	204.45	2577.4	1051.5	118517	3.15	24.58
1994	18653	29813.4	17042.1	261.53	3496.2	1357.1	119850	3.15	25.72
1995	19040	33070.5	20019.3	352.69	4283	1702.4	121121	1.98	26.86
1996	19922	36380.4	22913.6	430.29	4838.9	2024.2	122389	1.71	27.89
1997	20781	39762.7	24941.1	529.21	5160.3	2208.2	123626	1.44	28.29
1998	21616	42877.4	28406.2	615.56	5425.1	2336.7	124761	1.44	28.42
1999	22412	46144.6	29854.7	711.65	5854.02	2475.2	125786	0.99	28.32
2000	23151	50035.2	32917.7	823.37	6280	2694.7	126743	0.99	28.44
2001	23940	107450	37213.5	951.36	6859.6	2945.7	127627	0.99	28.61
2002	24780	117208	43499.9	1164.56	7702.8	3184.9	128453	0.72	28.72
2003	25639	128959	55566.6	1671.14	8472.2	3558	129227	0.72	29.32
2004	26476	141964	70477.4	2024.82	9421.6	4163	129988	0.72	30.72
2005	27331	156775	88773.6	2209.199	10493	5153	130756	0.72	31.96
2006	28310	204556	109998.2	2270.237	11759.5	5828	131448	0.72	33.35

资料来源：利率和汇率来源于 1991-2007《世界经济年鉴》，建筑业从业人员平均工资来源于 1991-2007《中国劳动统计年鉴》其它指标来源于 1991-2007 年的《中国统计年鉴》。

4.1.2 模型的参数估计

利用上节中建立模型的思路，我们先构建综合的差分方程组，接着用超定方程组来获得其最小二乘解，从而确定方程组的系数。为此，先对模型中的所有数据取对数。将取对数后的建筑业就业人数定义为 Z ，而将取对数后的建筑业总产值到城镇居民储蓄总

额定义为 Z1-Z8。然后用 Eviews 5.0 软件中的命令建立全国城镇就业的多变量自回归 VAR 模型。得到的 Z 方程组的解为如下：

$$z(t) = 0.5218z(t-1) + 0.009z_1(t-1) + 0.0033z_2(t-1) + 0.0251z_3(t-1) - 0.0665z_4(t-1) + 0.1014z_5(t-1) + 0.6387z_6(t-1) - 0.017z_7(t-1) - 0.113z_8(t-1) - 2.807$$

其中 F 值为 3644.288 R² 值为 0.9998817 残差和为 7.05×10⁻⁵

注：对于其它变量对应的方程，Pr>F 的概率均大于 0.01。

其余的方程回归系数见表 5。

表 5 方程组的系数矩阵

	Z	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8
Z (-1)	0.5218	2.788	-6.564	-1.929	-1.419	0.3626	-0.001	-19.53	-0.557
Z1(-1)	0.009	0.025	0.1266	0.1393	0.0556	0.0044	0.0001	0.0644	-0.011
Z2(-1)	0.0033	0.9989	1.7078	0.22	0.6557	0.5366	-0.002	3.2286	0.1713
Z3(-1)	0.0251	-0.824	0.4113	0.6337	-0.148	0.0312	0.0024	0.257	0.0769
Z4(-1)	-0.066	-0.162	-2.555	0.2315	0.1108	0.1347	0.0031	-7.313	-0.16
Z5(-1)	0.1014	2.9719	2.3732	0.2906	0.1546	0.3258	-0.003	3.7697	0.2919
Z6(-1)	0.6387	-1.144	16.758	9.4756	7.6754	-5.374	0.8959	52.801	-1.427
Z7(-1)	-0.017	0.0063	-0.039	-0.005	-0.034	-0.107	0.0005	-0.628	-0.008
Z8(-1)	-0.113	-13.23	-7.05	-3.664	-1.606	-0.526	0.0144	-8.882	-0.453
常数项	-2.807	13.844	-115.1	-85.17	-70.35	59.641	1.1932	-396.3	24.147

其中，Z，Z1-Z8 为取对数后的指标。Z (-1) 为全国城镇就业人数的一阶延迟。同理，Z8 (-1) 为 Z8 指标城镇化程度的一阶延迟。接着，通过多变量 VAR 模型得到的方程组的解，可以预测各指标的对数 Z，Z1-Z8。利用公式 exp (Z)，可以得到建筑业就业人数，以及 8 个主要指标 GDP 总值、国家总投资、教育经费、出口贸易额、市场化程度、人口总量、利率以及城镇化程度的预测值。

为了检验拟合的效果，我们采用 MATLAB7.0 编程来实现 1990-2007 年的所有实际数据与拟合数据的比较。拟合图如图 4 所示。

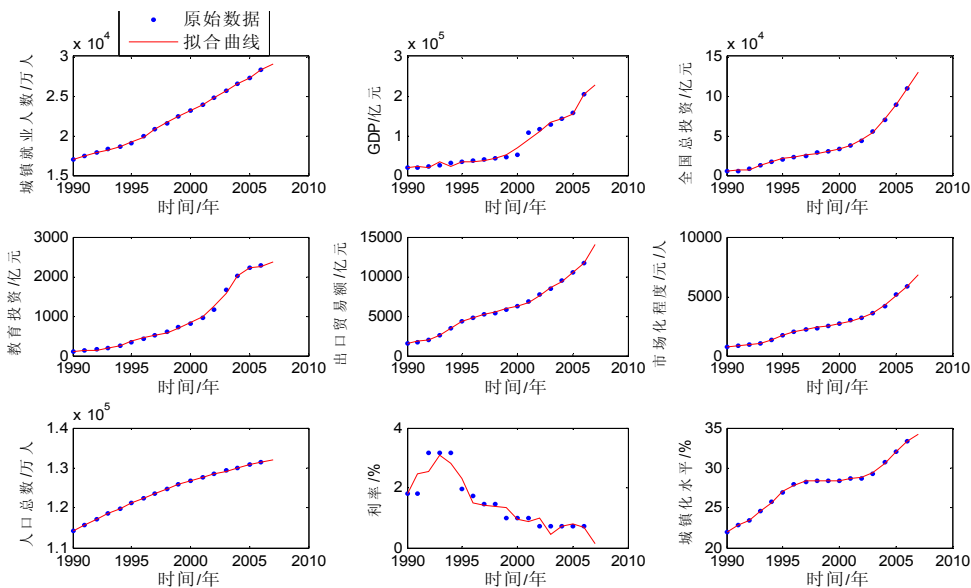


图 5 1990-2007 年各指标真实值与预测值的拟合情况比较

通过 MATLAB 计算获得的全国城镇就业人数，以及 8 个主要指标的实际值与预测值之间的误差率都较小，分别为 0.01%、0.46%、0.12%、0.23%、0.06%、0.08%、0%、0.89%以及 0.01%。故我们可以看出，通过建立建筑业的多变量自回归 VAR 模型，得到的方程组的解拟合效果非常好。

4.1.3 模型的检验

在 Eviews 软件中，点击 VAR 窗口工具栏中的 View 键会出现一个下拉菜单，其中一个 Lag Structure(滞后结构)，其中包含了 AR Boot Table(AR 根表)和 AR Boot Graph (AR 根图)。AR 根表是以列表形式给出 VAR 模型特征方程的根的倒数值，而 AR 根图以画图形式给出 VAR 模型特征方程的根的倒数值。该节中代入全国城镇就业人数及影响就业的各指标的数据，得到下图所示的 AR 根表和 AR 根图。

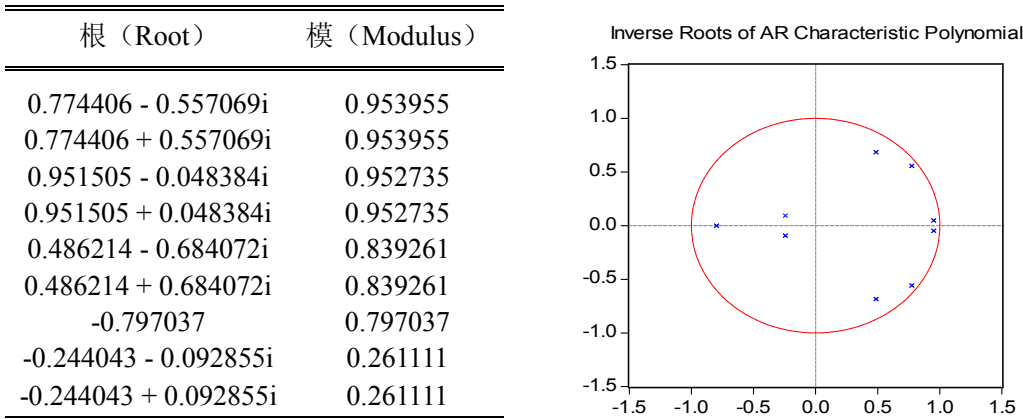


图 6 多变量 VAR 模型的稳定性检验

我们知道，如果模型中的根的模全部小于 1，则全部根的倒数值都在单位圆之内。而如果全部根的倒数值都在单位圆之内，则认为多变量 VAR 模型是稳定的。而从上图可以看出，全部根的倒数值都刚好落在单位圆之内，故该多变量 VAR 模型可以认为是稳定的。

接着对每个方程进行 Residual Tests (残差) 进行分析。结果表明，全国城镇就业人数，以及 8 个主要指标所建立的方程的残差都较小，而且其 R-squared 值以及矫正的 R-squared 都趋近于 1。故检验结果显示模型拟合较好。

依据上述分析，所做的残差图见图 6。

根据以上分析，可以发现，本文所建立的全国城镇就业人数及各影响因素的多变量 VAR 模型精度很高，拟合效果非常好。

4.1.4 模型的结论

由方程解得系数可以看出：城镇就业人数主要受去年全国人口总数，以及去年城镇就业人数的影响。去年的全国人口与城镇就业人数的系数即弹性是 0.6385，而去年就业人数对今年就业人数的影响系数是 0.5218.即若去年全国人口总数越大，今年城镇就业的总人数压力就会增大。而去年城镇就业人数的多少，也会影响今年的城镇就业人数。

GDP 总量主要受去年的市场化程度、城镇就业人数以及国家固定资产投资的影响较大影响系数分别是 2.9719、2.788 以及 0.9989.市场化程度越高，暗示居民的生活水平越高，人们就会有更多的钱投入生产消费，能够有助于 GDP 总量的增加。而去年城镇就业人数的多寡，一方面说明了去年就业市场的兴旺程度，另一方面也反映了去年城镇

就业对 GDP 总量的影响，也对今年的 GDP 总量有预警作用。

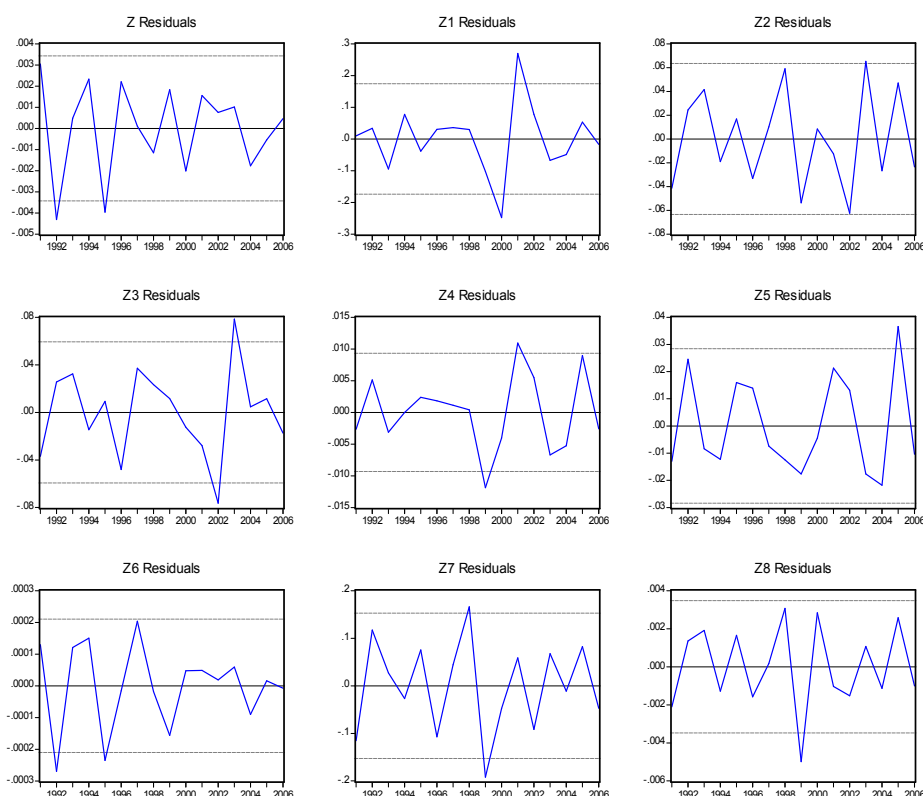


图 7 全国城镇就业人数及各影响因素的残差示意图

而众所周知，固定资产投资越多，市场和行业将吸收更多的劳动力就业，对 GDP 总量的提升有极大的促进作用。

国家固定资产投资主要受去年的城镇就业人数与全国总人数影响较大。国家的投资是有计划有比例的，去年的全国人口以及城镇就业人数都会影响今年的国家固定资产投资。而教育经费主要受去年的人口以及去年的教育经费影响。因为今年的教育经费一般是按照去年的人口总数以及去年的教育经费权衡计算获得的，所以这种影响是很明显的。

另外，出口贸易额以及市场化程度主要受由去年的国家总投资和去年的国家人口影响较大。去年的国家人口将极大地影响今年的人口，利率以及城镇化程度。

4.2 建立建筑行业城镇就业人数模型

本文力求分行业来建立确切的城镇就业人数模型。从《中国统计年鉴 2008》上可以看出，行业主要分为农业、工业、建筑业、运输和邮电业、社会服务业、旅游业、金融和保险业等行业。由于时间关系，给每个行业分别建立一个精确地城镇就业人数模型不太可能，工作量相当大。

众所周之，建筑行业是我国从业人数最多的行业之一，就业人数占全国城镇就业总人数的比重较大。同时，建筑业目前还是一个劳动密集型产业，是吸纳城镇就业人员和进城务工农民的主要行业。2006 年，建筑业就吸纳了约 1400 多万农村转移的劳动力。因此，本文选取建筑业为例，选取影响建筑业就业的指标，来建立城镇建筑行业就业人数与影响建筑业就业的指标的数学模型。

4.2.1 影响建筑行业的就业指标选取

影响建筑业就业的主要因素或指标很多。模型中的就业人数就改变为建筑业就业人数，指标则选取了 9 个主要指标，分别是建筑业总产值、国家对建筑业的总投资、出口贸易额、建筑业人员的平均工资、市场化程度、城镇化程度、人口总量、利率以及城镇居民的储蓄总额。通过查阅数据，得到了所有指标从 1990-2006 年的时间序列数据。如表 6 所示。

表 6 建筑业人数与各影响因素数据一览表

年份	建筑业 就业人 数	建筑业 总产值	国家对 建筑业 总投资	出口贸 易额	建筑业 平均工 资	市场化 程度	城镇化 程度	人口 总量	利率	城镇居民 储蓄总额
	万人	亿元	亿元	亿元	亿元	元	%	万人	%	亿元
1990	896	859.4	10.41	1510.2	2384	699.75	22.01	114333	1.8	7119.8
1991	940	1015.1	12.6	1700.6	2649	812.96	22.79	115823	1.8	9241.6
1992	995	1415	23.25	2026.6	3066	938.29	23.43	117171	3.15	11759.4
1993	1153	2266.46	115.02	2577.4	3779	1051.5	24.58	118517	3.15	15203.5
1994	1072	2964.69	138.35	3496.2	4894	1357.1	25.72	119850	3.15	21518.8
1995	1053	3728.85	145.55	4283	5785	1702.4	26.86	121121	1.98	29662.3
1996	1035	4387.35	183.36	4838.9	6249	2024.2	27.89	122389	1.71	38520.8
1997	1004	4621.61	150.95	5160.3	6655	2208.2	28.29	123626	1.44	46279.8
1998	846	4985.76	158.28	5425.1	7456	2336.7	28.42	124761	1.44	53407.5
1999	778.22	5172.1	223.65	5854	7982	2475.2	28.32	125786	0.99	59621.8
2000	744.2	5522.29	197.47	6280	8735	2694.7	28.44	126743	0.99	64332.4
2001	732.58	5931.67	189.24	6859.6	9484	2945.7	28.61	127627	0.99	73762.4
2002	756.08	6465.46	268.97	7702.8	10279	3184.9	28.72	128453	0.72	86910.6
2003	833.71	7490.78	290.06	8472.2	11478	3558	29.32	129227	0.72	103617
2004	841.03	8694.28	302.48	9421.6	12770	4163	30.72	129988	0.72	119555
2005	926.59	10133.8	351.12	10493	14338	5153	31.96	130756	0.72	141051
2006	988.67	11851.1	353.17	11760	16406	5828	33.35	131448	0.72	159768

资料来源：利率和汇率来源于 1991-2007《世界经济年鉴》，建筑业从业人员平均工资来源于 1991-2007《中国劳动统计年鉴》其它指标来源于 1991-2007 年的《中国统计年鉴》。

4.2.2 模型的参数估计

利用上节中建立模型的思路，我们先构建综合的差分方程组，接着用超定方程组来获得其最小二乘解，从而确定方程组的系数。为此，先对模型中的所有数据取对数。将取对数后的建筑业就业人数定义为 Z，而将取对数后的建筑业总产值到城镇居民储蓄总额定义为 Z1-Z9。然后运用 Eviews 5.0 软件建立建筑业的多变量自回归模型。拟合得到的方程组的参数值见表 7。

表 7 建筑业多变量自回归系数表

	Z	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
Z (-1)	0.8953	0.7443	0.351	0.3834	0.4551	0.6455	0.1623	-3E-04	1.985	0.4203
Z1(-1)	-0.721	-0.439	3.1813	-0.136	-0.591	-0.546	-0.025	0.0088	-2	0.0715

Z2(-1)	0.0963	0.0592	-0.206	0.033	0.0906	-0.006	0.0046	-8E-04	0.074	-0.083
Z3(-1)	0.3025	0.622	-1.797	0.513	-0.052	0.1165	-0.194	-8E-04	2.43	1.0676
Z4(-1)	0.0586	-0.003	-2.195	0.5612	1.0069	1.6329	0.3215	-0.004	-2.52	0.1348
Z5(-1)	1.5516	0.168	1.8764	-0.102	-0.103	-0.392	0.1051	0.0022	-2.09	-0.513
Z6(-1)	-3.864	-0.89	-18.15	-0.148	0.9614	2.6155	-0.02	-0.014	11.68	0.4186
Z7(-1)	-2.665	13.204	8.3487	11.323	12.634	9.192	-1.617	0.8887	61.82	14.142
Z8(-1)	0.229	0.3605	1.4179	0.0881	0.0798	-0.073	0.0097	-9E-04	0.278	-0.021
Z9(-1)	-0.106	0.2093	2.2415	-0.352	-0.175	-0.438	0.0823	-0.001	-1.63	-0.136
常数项	36.396	-154	-64.64	-130.1	-146.8	-116.2	18.57	1.3259	-726	-164.3

其中Z的方程为：

$$z(t) = 0.8953z(t-1) - 0.721z_1(t-1) + 0.0963z_2(t-1) + 0.3025z_3(t-1) + 0.0586z_4(t-1) + 1.5516z_5(t-1) - 3.864z_6(t-1) - 2.665z_7(t-1) + 0.229z_8(t-1) - 0.106z_9(t-1) + 36.396$$

其中F值为6.558179 R²值为0.92916 残差和为0.02211。

注：对于其它变量对应的方程，Pr>F的概率均小于0.01。

其中，Z，Z1-Z9为取对数后的指标。Z(-1)为建筑业就业人数的一阶延迟。同理，Z9(-1)为Z9指标城镇居民储蓄总额的一阶延迟。本文只列举了Z的方程式，其他指标的方程式以此类推。通过得到的方程组的解，可以预测各指标的对数Z，Z1-Z9。利用公式exp(Z)，可以得到建筑业就业人数，以及9个主要指标建筑业总产值、国家对建筑业的总投资、出口贸易额、建筑业人员的平均工资、市场化程度、城镇化程度、人口总量、利率以及城镇居民的储蓄总额的预测值。

为了检验拟合的效果，我们采用MATLAB7.0编程来实现1990-2006年的所有实际值与拟合值的比较。拟合图如图7所示。

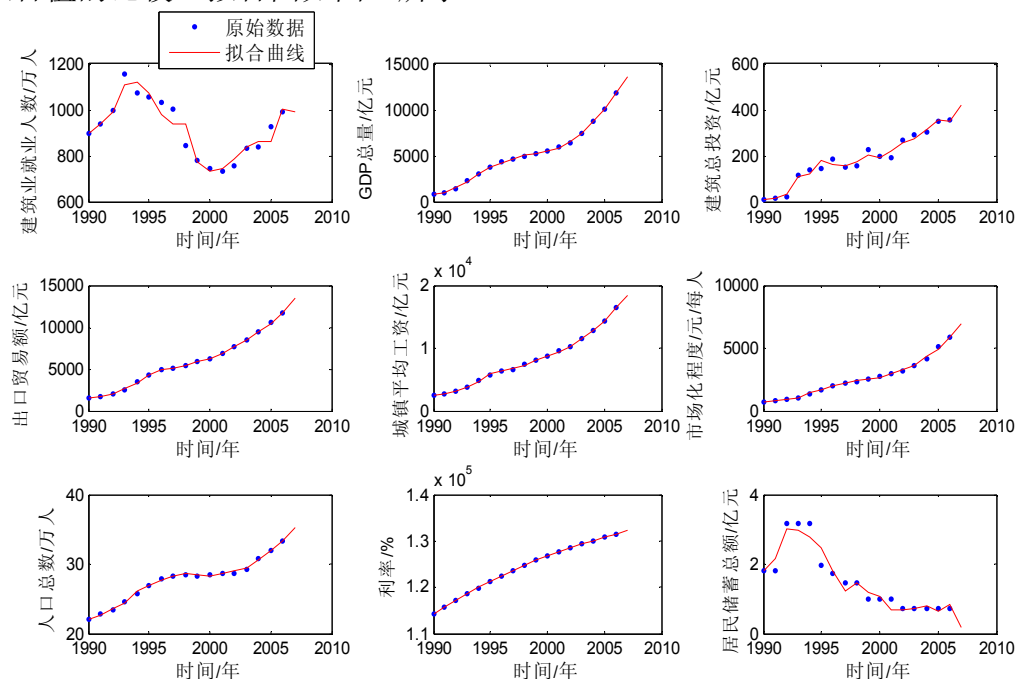


图8 1990-2006年各指标真实值与预测值的拟合情况比较

注：由于篇幅原因，城镇居民储蓄总额的拟合图被省略

通过 MATLAB 计算获得的建筑业就业人数，以及 9 个主要指标的误差率都较小。分别为 0.21%、0.07%、0.44%、0.04%、0.06%、0.14%、0.03%、0%、0.71%以及 0.07%。故我们可以看出，所建立得建筑业的多变量 VAR 模型，拟合效果不错。

4.2.3 模型的检验

在 Eviews 软件中，点击 VAR 窗口工具栏中的 View 键下的命令 Lag Structure（滞后结构），其中包含了 AR Boot Table（AR 根表）和 AR Boot Graph（AR 根图）。AR 根表是以列表形式给出 VAR 模型特征方程的根的倒数，而 AR 根图以画图形式给出 VAR 模型特征方程的根的倒数。该节中代入建筑业城镇就业人数及影响就业的各指标的数据，得到下图 8 所示的 AR 根表和 AR 根图。

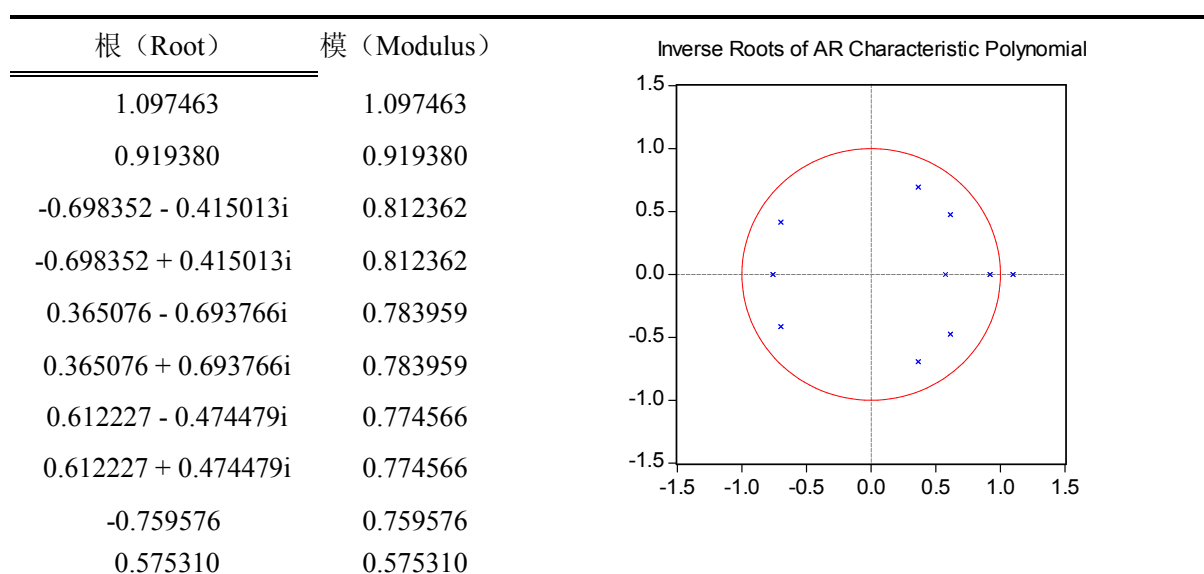


图 9 VAR 模型的稳定性检验

我们知道，如果模型中根的模全部小于 1，则全部根的倒数都在单位圆之内。而如果全部根的倒数都在单位圆之内，则认为 VAR 模型是稳定的，否则是非稳定的。而从上图可以看出，根的倒数中有一个落在单位圆之外，故该 VAR 模型是不稳定的。

于是就要对模型进行协整分析。本文使用 Johansen 检验，结果显示第 1 行检验通过，认为原假设成立，即 n 个变量之间存在协积关系。接着对每个方程进行 Residual Tests（残差）进行分析。结果表明，建筑业城镇就业人数，以及 8 个主要指标所建立的方程的残差都较小，而且其 R-squared 值以及矫正的 R-squared 都趋近于 1。故检验结果显示模型拟合较好。

依据上述分析，所做的残差图见表 9。

根据以上分析，可以发现，本文所建立的建筑业就业人数及各影响因素的多变量自回归 VAR 模型拟合效果较好。

4.2.4 模型的结论

由以上所得的系数矩阵可以看出，今年的建筑业就业人数主要由去年的市场化程度以及去年的建筑业就业人数决定。其中，去年市场化程度对今年建筑业就业人数的相关系数是 1.5516，而去年的建筑业就业人数也将极大地影响今年的建筑业就业人数。

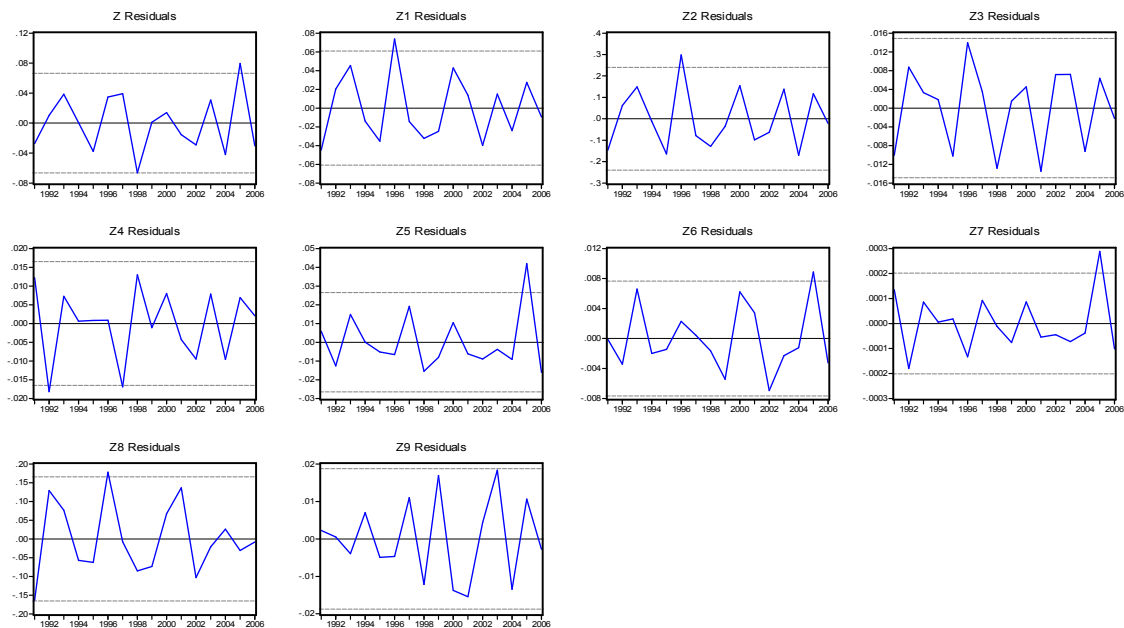


图 10 建筑业城镇就业人数及各影响因素的残差示意图

建筑业总产值主要由去年的全国人口总数决定。去年的全国人口越多，城镇就业人数也会相应增多，而建筑业是吸纳城镇就业人员最多的行业之一，故结果与实际情况较为吻合。

而出口贸易额、建筑行业平均工资、市场化程度、利率以及城镇居民储蓄总额都主要由去年的人口总量决定。说明人口总量的大小，影响了城镇就业人数的多寡，也对从事建筑行业的城镇劳动力产生了很大的影响。其中，城镇化程度由去年的建筑业就业人数主要决定，说明了建筑业就业人数的多寡在一定程度上影响了城镇化程度的高低。

4.3 建立浙江地区的城镇就业人数模型

本文力求分地区来建立确切的城镇就业人数模型。从《中国统计年鉴 2008》上可以看出，全国共有 31 个省（市、自治区）。由于浙江省在城镇就业方面做得比较好，也非常具有典型性，因此本文选取浙江省为例，来分析影响地区城镇就业的指标，来建立浙江省城镇就业人数与影响浙江省就业指标的数学模型。

4.3.1 影响浙江地区就业的指标选取

影响浙江地区就业的主要因素或指标很多。模型中的就业人数就变为浙江省的城镇就业人数，指标分别是浙江省的 GDP、国家对浙江省的固定资产投资、劳动力教育、出口贸易额、城镇人均可支配收入、利率、市场化水平、浙江省人口总数、城镇化水平以及城镇居民储蓄总额。通过查阅数据，得到了所有指标从 1990-2006 年的时间序列数据。如表 8 所示。

表 8 浙江省城镇就业人数与各影响因素数据一览表

年份	浙江省 就业人 数	浙江省 GDP	国家对 浙江省 总投资	劳动力 教育	出口贸 易额	城镇人 均可支 配收入	利 率	市场化 水平	浙江省人 口 总量	城镇化 水平	城镇居 民储蓄 总额
	万人	亿元	亿元	亿元	亿美元	元	%	元/人	万人	%	亿元

1990	505.86	897.99	186.96	4.12	20.19	1233.12	1.80	353.75	4168.00	26.41	606.01
1991	530.14	1131.50	239.75	4.34	27.98	1391.76	1.80	404.00	4202.00	28.90	789.64
1992	535.34	1365.06	361.18	5.82	33.52	1656.48	3.15	493.87	4236.00	31.00	1036.72
1993	563.23	1909.49	683.83	10.13	43.21	2315.76	3.15	493.06	4266.00	32.70	1316.53
1994	520.36	2666.86	1006.39	13.75	60.85	3291.24	3.15	695.06	4294.00	33.10	1910.98
1995	529.51	3524.79	1357.90	19.27	76.93	3891.00	1.98	1472.66	4319.00	35.40	2623.60
1996	543.47	4146.06	1617.53	27.09	80.41	4342.08	1.71	1599.40	4343.00	37.90	3400.19
1997	546.02	4638.24	1694.57	31.10	100.85	4824.12	1.44	1757.17	4435.00	40.10	4297.07
1998	550.79	4987.50	1847.93	40.98	108.63	5070.12	1.44	1909.00	4456.00	42.00	5264.21
1999	565.02	5364.89	1886.04	49.91	128.71	5472.48	0.99	2075.75	4475.00	45.10	6273.15
2000	583.83	6036.34	2267.22	63.78	204.82	6012.84	0.99	2553.59	4596.00	48.70	7299.57
2001	623.02	6748.15	2776.69	108.71	242.60	7073.76	0.99	2684.90	4613.00	50.00	8823.12
2002	672.69	7796.00	3596.31	110.30	315.65	7968.12	0.72	2847.70	4647.00	51.90	11242.84
2003	750.00	9395.00	4993.57	112.83	415.95	9570.48	0.72	3157.10	4680.00	53.00	14758.15
2004	847.67	11648.70	6059.78	117.60	581.39	12197.04	0.72	4055.50	4720.00	54.00	17236.62
2005	904.34	11648.70	6696.25	124.30	768.02	15743.16	0.72	4631.70	4898.00	56.00	20494.16
2006	1077.89	15742.51	7593.66	135.80	1008.91	19102.08	0.72	5325.30	4980.00	56.50	23362.16

资料来源：利率和汇率来源于 1991-2007《世界经济年鉴》，城镇人均可支配收入来源于 1991-2007《中国劳动统计年鉴》其它指标来源于 1991-2007 年的《中国统计年鉴》。

4.3.2 模型的参数估计

利用上节中建立模型的思路，我们先构建综合的差分方程组，接着用超定方程组来获得其最小二乘解，从而确定方程组的系数。为此，先对模型中的所有数据取对数。将取对数后的浙江省城镇就业人数定义为 Z，而将取对数后的浙江省 GDP 到城镇居民储蓄总额定义为 Z1-Z10。然后用 Eviews 5.0 软件中的命令来建立建筑业的多变量 VAR 模型。得到的方程组的回归参数拟合值矩阵见表 9。

表 9 浙江省城镇就业与其影响因素的多变量回归系数表

	Z	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
Z (-1)	0.5237	-0.698	-0.696	-1.977	0.3211	0.5048	0.5298	-2.189	0.115	-0.081	0.0338
Z1(-1)	-0.85	-1.242	-1.868	-2.219	-3.872	-1.31	3.5444	0.0381	-0.12	-0.111	0.6308
Z2(-1)	0.1179	0.8292	1.1537	0.5562	0.2564	0.505	0.104	-0.368	-0.07	-0.189	0.0623
Z3(-1)	0.0004	-0.104	-0.147	-0.28	-0.312	-0.243	-0.626	-0.19	-0.03	0.0043	0.2196
Z4(-1)	0.3056	0.1876	0.8828	0.0131	0.2771	-0.098	0.4529	0.7676	-0.04	0.0428	0.2033
Z5(-1)	-0.095	0.3842	0.1488	2.0066	3.8059	1.4906	-2.478	1.6396	0.29	0.2585	-0.443
Z6(-1)	0.0238	-0.013	0.2366	0.135	-0.2	0.0748	0.1776	-0.198	-0	0.0239	-0.032
Z7(-1)	0.113	-0.015	0.0193	0.4516	-0.027	0.1081	-0.485	-0.401	0.018	0.018	-0.192
Z8(-1)	-2.043	-4.389	-14.47	4.6584	-13.22	-1.079	-5.577	-10.9	-0.4	-2.305	-0.555
Z9(-1)	-0.807	1.2114	2.9397	6.1029	5.4877	2.332	3.7956	0.0256	0.409	0.7846	-0.114
Z10(-1)	0.5379	0.458	0.4478	-0.447	0.3538	-0.07	-1.511	0.3589	-0.03	0.1781	0.4726
常数项	20.667	40.389	112.72	-46.77	95.703	2.4264	32.148	91.455	9.289	18.623	5.325

其中Z的方程为

$$z(t) = 0.5237z(t-1) - 0.85z_1(t-1) + 0.1179z_2(t-1) + 0.0004z_3(t-1) + 0.3056z_4(t-1) + 0.095z_5(t-1) + 0.0238z_6(t-1) + 0.113z_7(t-1) - 2.043z_8(t-1) - 0.807z_9(t-1) + 0.5379z_{10}(t-1) + 20.67$$

其中 F 值为 184.6858 R^2 值为 0.998525 残差和为 0.000661。

注：对于其它变量对应的方程， $Pr > F$ 的概率均大于 0.01。

其中，Z，Z1-Z10 为取对数后的指标。Z(-1) 为浙江省城镇就业人数的一阶延迟。同理，Z10(-1) 为 Z10 指标城镇居民储蓄总额的一阶延迟。接着，通过多变量 VAR 模型得到的方程组的解，可以预测各指标的对数 Z，Z1-Z9。利用公式 $\exp(Z)$ ，可以得到浙江省的 GDP、国家对浙江省的固定资产投资、劳动力教育、出口贸易额、城镇人均可支配收入、利率、市场化水平、浙江省人口总数、城镇化水平以及城镇居民储蓄总额的预测值。

为了检验拟合的效果，我们采用 MATLAB7.0 编程来实现 1990-2006 年的所有实际值与拟合值的比较。拟合图如图 10 所示。

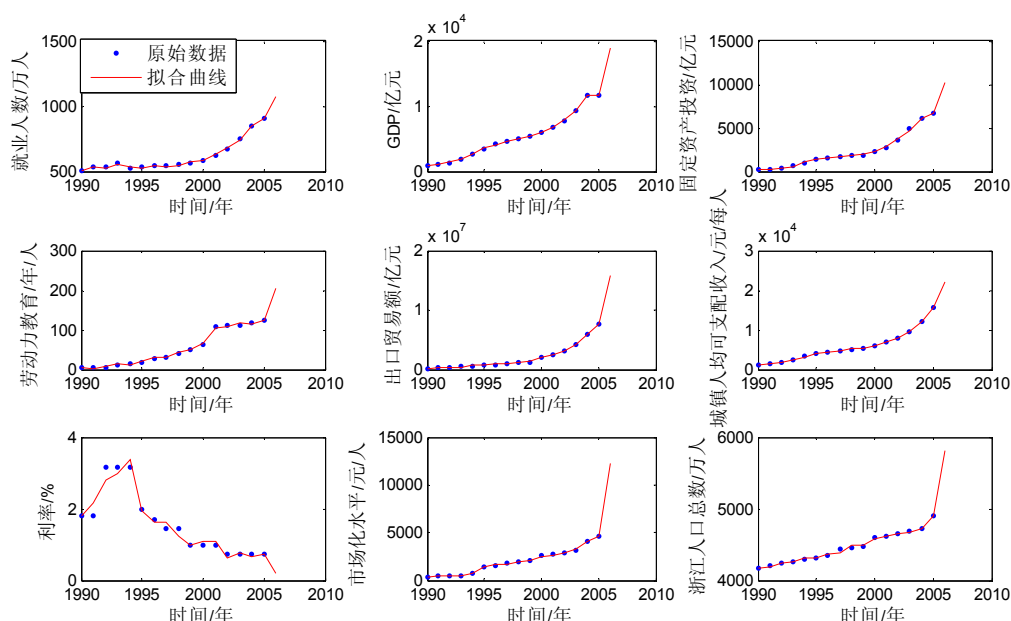


图 11 1990-2006 年各指标真实值与预测值的拟合情况比较

注：由于篇幅原因，城镇化水平和城镇居民储蓄总额的拟合图被省略。

通过 MATLAB 计算获得的浙江省城镇就业人数，以及 9 个主要指标的误差率都较小，分别为 0.07%、0.05%、0.23%、0.23%、0.13%、0.05%、0.57%、0.11%、0.02%、0.03% 以及 0.13%。故我们可以看出，所建立的浙江省城镇就业人数的多变量 VAR 模型，得到的方程组的解拟合效果很好。

4.3.3 模型的检验

在 Eviews 软件中，View 键下的菜单 Lag Structure(滞后结构)，其中包含了 AR Boot Table (AR 根表) 和 AR Boot Graph (AR 根图)。AR 根表是以列表形式给出 VAR 模型特征方程的根的倒数值，而 AR 根图以画图形式给出 VAR 模型特征方程的根的倒数值。该节中代入浙江省城镇就业人数及影响就业的各指标的数据，得到如图 11 所示的 AR 根表和 AR 根图。

根 (Root)	模 (Modulus)
1.343510	1.343510
0.857407 - 0.689215i	1.100074
0.857407 + 0.689215i	1.100074
0.529992 - 0.808318i	0.966576
0.529992 + 0.808318i	0.966576
0.954928	0.954928
-0.916772	0.916772
-0.576578 - 0.601196i	0.832994
-0.576578 + 0.601196i	0.832994
-0.221163 - 0.761561i	0.793025
-0.221163 + 0.761561i	0.793025

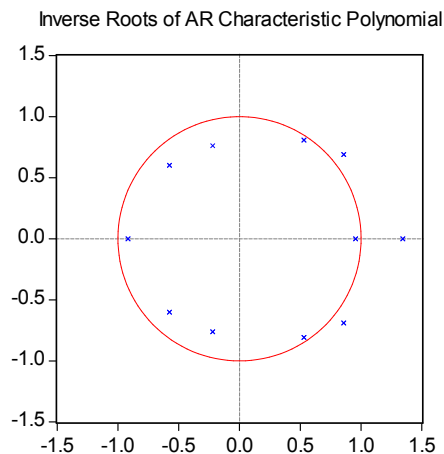


图 12 VAR 模型的稳定性检验

我们知道，如果模型中根的模全部小于 1，则全部根的倒数都在单位圆之内。而如果全部根的倒数都在单位圆之内，则认为 VAR 模型是稳定的，否则是非稳定的。而从上图可以看出，根的倒数有 3 个落在单位圆之外，故该 VAR 模型是不稳定的。

于是就要对模型进行协整分析。本文使用 Johansen 检验，结果显示第 1 行检验通过，认为原假设成立，即 n 个变量之间存在协积关系。接着对每个方程进行 Residual Tests（残差）进行分析。结果表明，建筑业城镇就业人数，以及 8 个主要指标所建立的方程的残差都较小，而且其 R-squared 值以及矫正的 R-squared 都趋近于 1。故检验结果显示模型拟合较好。依据上述分析，所做的残差图如图 12 所示：

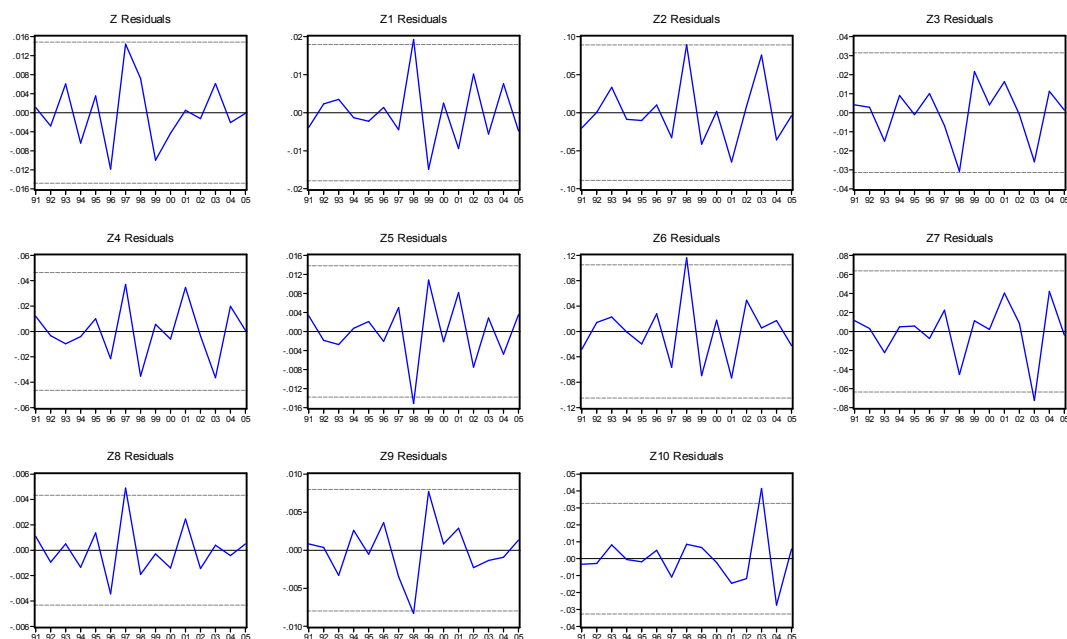


图 13 全国城镇就业人数及各影响因素的残差示意图

根据以上分析，可以发现，本文所建立的全国城镇就业人数及各影响因素的高维 VAR 模型精度很高，拟合效果非常好。

4.3.4 模型的结论

由解得的方程系数矩阵，可以看出，今年浙江省的就业人数主要由前一年的浙江省就业人数以及城镇居民储蓄总额决定。一般而言，前一年对后一年的就业人数有一定的参考作用。去年的城镇居民储蓄总额在一定程度上反映了浙江省的城镇居民的生活水平。生活水平越高，城镇就业人数就会越多。

浙江省 GDP 主要受前一年国家对建筑业的总投资以及城镇化水平决定。国家对浙江省的固定资产投资、劳动力素质、出口贸易额、城镇人均可支配收入以及城镇化水平都受前一年城镇化水平的影响较大。其中，出口贸易额还受到城镇人均可支配收入的影响，市场化水平主要受到城镇人均可支配收入的影响。

4.4 建立农民工群体的城镇就业人数模型

本文力求分群体来建立确切的城镇就业人数模型。从群体角度看，影响城镇就业最明显的群体是农民工群体。农民工，是指在本地乡镇企业或者进入城镇务工的农业户口人员。农民工是我国特有的城乡二元体制的产物，是我国在特殊的历史时期出现的一个特殊的社会群体。农民工有广义和狭义之分：广义的农民工包括两部分人，一部分是在本地乡镇企业就业的离土不离乡的农村劳动力，一部分是外出进入城镇从事二、三产业的离土又离乡的农村劳动力；狭义的农民工主要是指后一部分人。据有关部门的调查，我国狭义农民工的数量为 1.2 亿人左右，广义农民工的数量大约为 2 亿人。本文选取农民工群体为例，选取影响农民工城镇就业的指标，来建立农民工城镇就业人数与影响就业指标的数学模型。

4.4.1 影响农民工群体就业的指标选取

影响农民工群体就业的主要因素或指标很多。模型中的就业人数用农村劳动力外出人数替代，因为农村劳动力的绝大部分进入到城镇就业，因此，本文用农村劳动力外出人数来近似替代农民工群体的城镇就业人数。由于影响农民工群体就业的因素或指标非常多，在此，只选取了一些主要因素。它们分别是市场化水平、城镇化程度、城乡收入差距、出口额、二、三 GDP 总额以及人口总数。通过查阅数据，得到了所有指标从 1990-2007 年的时间序列数据。如表 10 所示。

表 10 农民工城镇就业人数与各影响因素数据一览表

年	外出打工人数	市场化程度	城镇化程度	城乡收入差距	出口额	二三 GDP 总额	人口总数
份	万人	元/人	%	元	亿元	亿元	万人
1990	1800	699.75	22.01	823.9	2985.8	3809.7	114333
1991	2140	812.96	22.79	992	3827.1	4819.1	115823
1992	2592	938.29	23.43	1242.6	4676.3	5918.9	117171
1993	2752	1051.5	24.58	1655.8	5284.8	6940.6	118517
1994	2888	1357.1	25.72	2275.2	10421.8	12697	119850
1995	3000	1702.4	26.86	2705.3	12451.8	15157.1	121121
1996	3400	2024.2	27.89	2912.8	12576.4	15489.2	122389
1997	3890	2208.2	28.29	3070.2	15160.7	18230.9	123626
1998	4936	2336.7	28.42	3263.1	15223.6	18486.7	124761
1999	5240	2475.2	28.32	3643.72	16159.8	19803.52	125786
2000	7600	2694.7	28.44	4026.6	20634.4	24661	126743
2001	9050	2945.7	28.61	4493.2	22024.4	26517.6	127627

2002	10470	3184.9	28.72	5227.2	26947.9	32175.1	128453
2003	11390	3558	29.32	5850	36287.9	42137.9	129227
2004	11823	4163	30.72	6485.2	49103.3	55588.5	129988
2005	12578	5153	31.96	7238.1	62648.1	69886.2	130756
2006	13283	5828	33.35	8172.5	77594.59029	85767.09029	131448
2007	13988	6769	34.47	9645.4	93455.62719	103101.0272	132129

资料来源：城乡收入差距来源于 1991-2008《中国劳动统计年鉴》；其它指标来源于 1991-2008 年的《中国统计年鉴》。

4.4.2 模型的参数估计

利用上节中建立模型的思路，我们先构建综合的差分方程组，接着用超定方程组来获得其最小二乘解，从而确定方程组的系数。为此，先对模型中的所有数据取对数。将取对数后的农民工城镇就业人数定义为 Z，而将取对数后的市场化程度到人口总数定义为 Z1-Z6。然后用 Eviews 5.0 软件的命令来建立农民工就业的多变量自回归模型。得到的方程组的回归参数拟合值矩阵见表 11。

表 11 浙江省城镇就业与其影响因素的多变量回归系数表

	Z	Z1	Z2	Z3	Z4	Z5	Z6
Z (-1)	-0.168	-0.133	-0.033	0.3424	-0.386	-0.263	-7E-04
Z1(-1)	0.8246	0.2158	0.0003	-0.951	-0.009	-0.146	0.002
Z2(-1)	-6.188	-0.682	0.2636	2.7235	-5.338	-4.085	0.0011
Z3(-1)	-1.52	-0.676	-0.16	0.1727	-2.875	-2.485	0.0023
Z4(-1)	-4.974	-4.752	-1.317	-4.138	-21.21	-18.77	-0.003
Z5(-1)	6.0638	5.9919	1.6397	4.9553	25.878	22.911	0.0019
Z6(-1)	27.157	4.3511	-0.583	0.6071	9.7184	8.0907	0.9033
常数项	-293.6	-49.14	7.4447	-13.8	-109.9	-92.24	1.1178

其中Z的方程为：

$$z(t) = -0.168z(t-1) + 0.825z_1(t-1) - 6.1879z_2(t-1) - 1.52z_3(t-1) - 4.97z_4(t-1) + 6.064z_5(t-1) + 27.157z_6(t-1) - 293.62$$

其中F值为184.6858 R²值为0.998525 残差和为0.000661。

注：对于其它变量对应的方程，Pr>F 的概率均大于 0.01。

其中，Z，Z1-Z6 为取对数后的指标。Z（-1）为农民工城镇就业人数的一阶延迟。同理，Z6（-1）为 Z10 指标人口总数的一阶延迟。通过方程组解的矩阵，可以预测各指标的对数 Z，Z1-Z6。利用公式 $\exp(Z)$ ，可以得到市场化水平、城镇化程度、城乡收入差距、出口额、第二三产业 GDP 总额以及人口总数的预测值。

为了检验拟合的效果，我们采用 MATLAB7.0 编程来实现 1990-2007 年的所有实际值与拟合值进行比较。拟合图如图 13 所示。

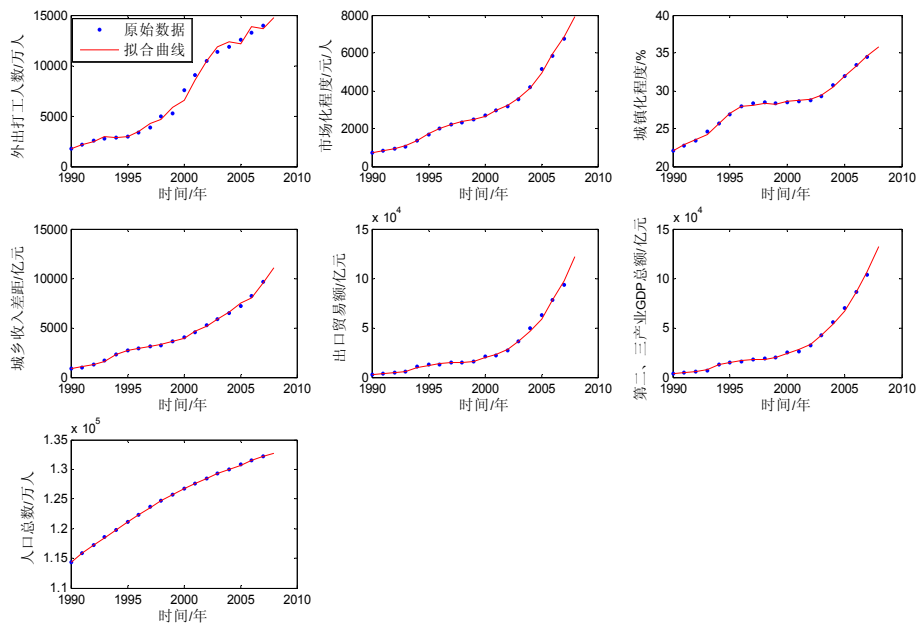


图 14 1990-2007 年各指标真实值与预测值的拟合情况比较

通过 MATLAB 计算获得的农民工城镇就业人数，以及 6 个主要指标的误差率都较小，分别为 0.27%、0.08%、0.02%、0.09%、0.22%、0.18%以及 0%。故我们可以看出，通过建立农民工就业的多变量回归模型，得到的方程组的解拟合效果非常好。

4.4.3 模型的检验

在 Eviews 软件中，点击 View 键下的 Lag Structure(滞后结构)，其中包含了 AR Boot Table (AR 根表) 和 AR Boot Graph (AR 根图)。AR 根表是以列表形式给出 VAR 模型特征方程的根的倒数，而 AR 根图以画图形式给出 VAR 模型特征方程的根的倒数。该节中代入全国城镇就业人数及影响就业的各指标的数据，得到下图 14 所示的 AR 根表和 AR 根图。

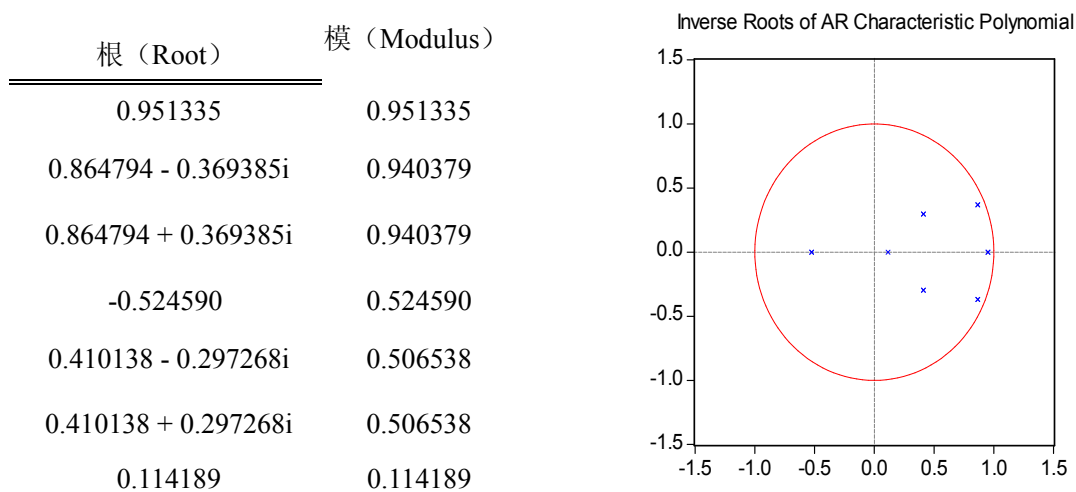


图 15 VAR 模型的稳定性检验

我们知道，如果模型中根的模全部小于 1，则全部根的倒数都在单位圆之内。而如果全部根的倒数都在单位圆之内，则认为 VAR 模型是稳定的，否则是非稳定的。

而从上图可以看出，全部根的倒数都落在单位圆之内，故该 VAR 模型是稳定的。

接着对每个方程进行 Residual Tests（残差）进行分析。结果表明，农民工就业人数，以及 8 个主要指标所建立的方程的残差都较小，而且其 R-squared 值以及矫正的 R-squared 都趋近于 1。故检验结果显示模型拟合较好。

接着对该模型关于就业率总人数做脉冲响应函数分析。依据上述分析，所做脉冲图如图 15 所示：

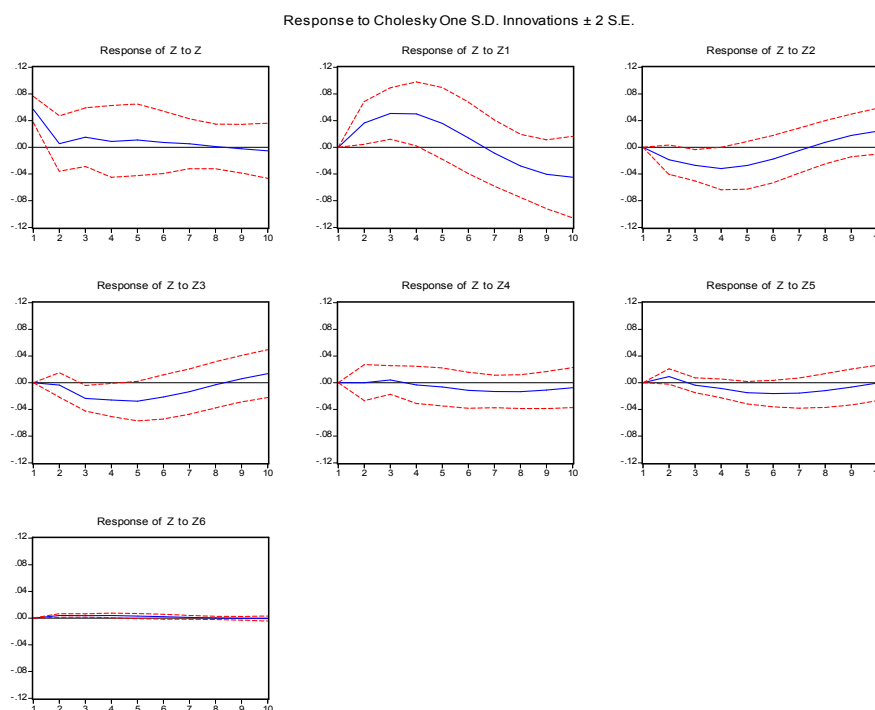


图 16 农民工就业人数因素的脉冲图

上图的 7 个子图代表 7 个指标在受到某种冲击时对系统的动态影响。由上图可见其中子图 5、6、7 对应的指标：出口额、第二、三产业 GDP 总额。在受到一个冲击后对系统的冲击幅度不是很大。而子图 2、3、4 对应的指标：市场化水平、城镇化程度、城乡收入差中对应的指标在受到一个冲击后对系统会有较长的持续效应。因此，可见对于农民工就业问题主要需要考虑对就业总数影响较大的指标。政府可以利用这个结论，对就业问题进行有区别有重点的采取措施。

4.4.4 模型的结论

以农民工为主体的城市外来务工人员在社会做出重大贡献的同时，一直遭到职业歧视和待遇歧视的双重不公平对待。农民工就业机会虽多，但是由于与城市居民的受教育程度存在差异，农民工就业的工作往往比较差，劳动时间长，劳动强度大，工作危险而且往往不能得到应有的报酬，农民工在就业方面受到很多限制和歧视。因此，本文对影响农村工城镇就业因素的研究，具有较好的现实意义。

由解得的方程系数矩阵，可以看出，农民工群体外出打工人数主要由农民工人口总数决定。农民工人口总数越大，有更多的农民工倾向于外出打工。而市场化的程度主要受第二三产业 GDP 总额以及人口总数的影响。而城镇化程度、城乡收入差距出口额以及第二三产业总额都受人口总数的影响较大。前一年的人口总数也对后一年的人口总数有影响。并且，通过多次对向量自回归模型的拟合，显示了该模型具有较强的稳定性和可操作性，你和效果好，预测精度高。

5 问题四解答

——仿真 2009 年及 2010 年上半年的就业人数

5.1 考虑经济危机的仿真方案

该仿真模型主要考虑在经济危机影响下，问题二中模型的变量值受到经济危机的影响会发生突变，从而会影响总的就业人数。本文假设经济危机只会给社会总投资额、出口贸易额带来较大的影响，同时，考虑国家在 2009 年投资的 4 万亿带来的投资增加额。其他指标的数值本文通过时间序列分析进行预测，得到其他指标的一个取值区间，认为 2008 年度以后指标的取值均在以上求得的区间中。

时间序列分析得到各个指标置信度为 95% 的可行区间，见表 12。

表 12 指标预测区间

指标	置信区间	模型
GDP 总额	[141552,256675]	AR(1,1)
教育经费	[1696,2771]	AR(1,1)
人口总量	[129814,134300]	AR(1,1)
市场程度	[5798,7662]	AR(1,1)
城镇化水平	[30.6,35.7]	AR(1,1)

注：以上预测模型参数均通过检验，结果给出了预测区间。

比较资讯网 2009 年上半年与 2008 年上半年逐月出口贸易额同比数据，发现，从 1 月份至 6 月份，每月出口量同比下降 20% 左右。于是我们可假定，2009 年出口量比 2008 年下降 18-22%。根据 2008 年我国出口量为 14291.7 亿元的数据可以计算，2009 年我国出口贸易额为 [11148,11719] 亿元。

最后，根据以上分析得到 2008 年的各指标的置信区间，仿真过程中在区间中随机取值，代入问题二中得到模型中即可以预测 2009 年总的就业人数。经过多次随机模拟可以得到的到一系列不同的就业人口总数，然后分析所得数据，可以得到 2009、2010 年的就业人数服从正态分布见图 17、18。

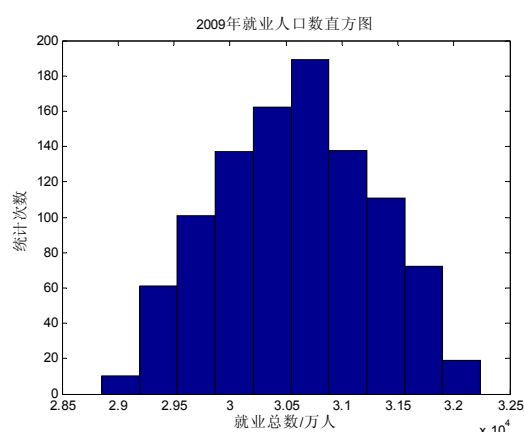


图 17 2009 年城镇就业总人数分布函数图

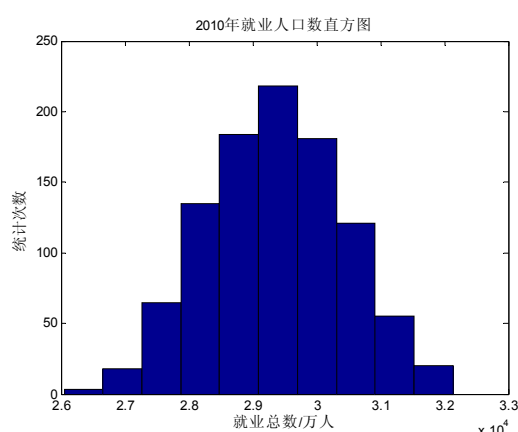


图 18 2010 年城镇就业总人数分布函数图

多次随机模拟得到的图像基本如上图所示。对以上数据做正态分布检验，本文用偏度-峰度检验法得到以上两年的数据检验结果见表 13。

表 13 指标预测区间

检验结果	峰度值	偏度值	偏态	均值	方差
2009 年	2.31	-0.013	左偏态	3.058×10^4	0.0697×10^4
2010 年	2.63	0.031	右偏态	2.935×10^4	0.106×10^4

由此可知,两年就业人数显然都服从正态分布,2009 年的城镇就业总人数在置信度为 95%的置信区间为 $[2.925 \times 10^4, 3.175 \times 10^4]$ (单位:万人),2010 年的总人数在置信度为 95%的置信区间为 $[2.75 \times 10^4, 3.15 \times 10^4]$ (单位:万人)。就业人数下降明显,可见经济危机对我国城镇就业的影响较大。

由于考虑影响因素还不能充分反映出经济危机给就业带来的影响,仿真预测的总人数存在误差较大。由此,本文在问题二的模型上进行改进。

5.2 考虑国外投资特殊性的仿真方案二

5.2.1 模型二的分析

吸引FDI (Foreign Direct Investment, 对外直接投资)有利于解决国内资金短缺和“双缺口”问题,促进产品的更新与产业结构的升级,还可以学习和引进国际上先进的技术水平和管理经验。随着中国逐步融入世界经济体系,FDI对我国就业人数影响越来越大,积极吸引FDI目前已成为我国增加就业人数的一个重要政策。

我国劳动力市场的最大特点是劳动力廉价。据相关统计资料(《FDI对中国就业影响的动态分析》^[26]),外商投资在劳动密集型的产业占有很大一部分比例,吸引大量人员就业。

金融危机爆发后,流入我国的FDI显著减少,据相关报道(<http://www.p5w.cn/news/gncj/200908/t2515362.htm>),截止2009年7月份的实际使用外资数据已连续10个月负增长。今年7月,我国实际使用外资金额53.59亿美元,同比下降35.71%。与此同时,据调查(<http://finance.sina.com.cn/g/20090302/00315915579.shtml>),仅农民工群体,就有2000万农民工受金融危机影响失业而返乡,对社会稳定和发展造成很大压力。

本题在金融危机的大背景下,将投资(Investment)分为FDI和DI(Domestic Investment, 本国投资),通过研究两种不同投资对我国就业的不同影响力,得出有效影响就业的有效就业投资 G' 。

5.2.2 模型二的建立

总投资 G 可以分为国外投资FDI和国内投资DI, 即:

$$G = 1 \cdot FDI + 1 \cdot DI$$

联系实际,总投资 G 与就业人数往往没有紧密的关系,那是因为FDI和DI投资产业和方向不同,所以FDI和DI对就业人数的贡献度是不同的。一般情况下,FDI对增加就业人数的贡献度要大于DI的贡献度,所以导致总投资 G 变化的效果往往不能有效反映在就业人数的变动上。本文通过定量的计算,得出两种投资对就业人数的贡献系数,进而计算出和就业人数关系密切的有效就业投资:

$$G' = k_1 \cdot FDI + k_2 \cdot DI \quad (k_1 + k_2 = 2)$$

首先定量确定就业人数和FDI和DI的关系,我们以就业人数作为因变量,以FDI和DI作为自变量,建立二元线性回归模型:

$$Y = \alpha X_1 + \beta X_2 + \varepsilon$$

符号说明: Y ——我国就业人数 X_1 ——FDI X_2 ——DI

首先,收集二元线性回归模型所需要用到的数据,见表14。

表14 求解模型所收集的数据

年份	城镇就业人员（万人）	国外投资（亿元）	国内投资（亿元）
1990	17041	492.1434	16548.86
1991	17465	552.6509	16912.35
1992	17861	918.5179	16942.48
1993	18262	1863.535	16398.47
1994	18653	2066.964	16586.04
1995	19040	4019.587	15020.41
1996	19922	4556.597	15365.4
1997	20781	5339.294	15441.71
1998	21616	4847.993	16768.01
1999	22412	4359.27	18052.73
2000	23151	4913.727	18237.27
2001	23940	4111.351	19828.65
2002	24780	4553.26	20226.74
2003	25639	4646.708	20992.29
2004	26476	5303.111	21172.89
2005	27331	5226.714	22104.29
2006	28310	5347.165	22962.84
2007	29350	5956.898	23393.1

注：数据来源于《2008中国统计年鉴》和《2008中国经济统计年鉴》

通过SAS计算，结果如下：

表15 二元回归分析表

变量	系数	标准差	t 值	Pr> t
x_1	0.78613	0.22077	3.56	0.0028
x_2	0.07502	0.01096	6.85	<0.001
常数	16479	708.56229	23.26	<0.001

上表显示， $\text{Pr}>|t|$ 值<0.05,三个参数均通过检验：

则模型如下：

$$Y = 0.78613X_1 + 0.07502X_2 + 16479$$

可以看出，FDI的系数 α 为0.78613远大于DI系数0.07502，说明FDI在吸引就业人数上要明显大于DI。

通过解方程：

$$\begin{cases} \frac{\alpha}{\beta} = \frac{k_1}{k_2} \\ k_1 + k_2 = 2 \end{cases}$$

可以得出 $k_1 = 1.826$ $k_2 = 0.154$ 即有效就业投资： $G' = 1.826 \cdot FDI + 0.154 \cdot DI$

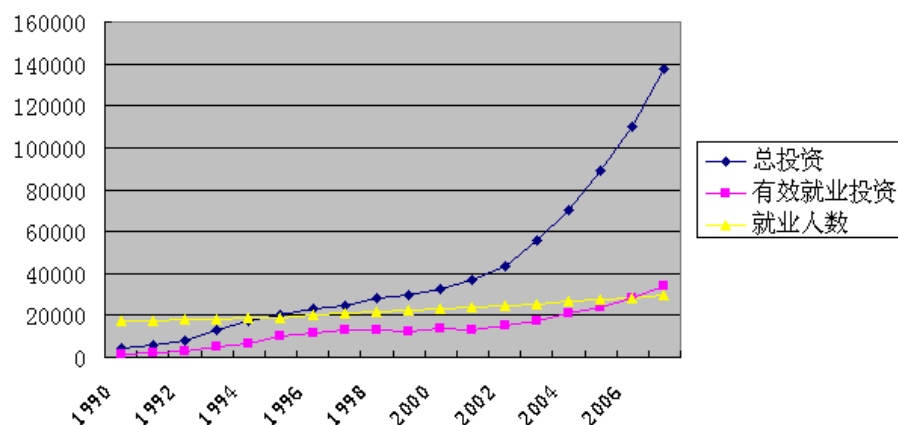


图19 投资与就业人数变化图

通过图像，可以明显看出，随着时间的推移，总投资 G 的增长速度明显大于就业人数的增加速度，而有效就业投资 G' 增加速度与就业人口增加速度基本相同，与就业人数的发展有很强的同向性。

为了进一步验证我们的结论，我们对总投资 G ，有效就业投资 G' 和就业人数进行相关系数检验，结果如下：

表16 三变量相关系数

	就业人数	有效就业投资	总投资
就业人数	1	--	--
有效就业投资	0.9578	1	--
总投资	0.9295	0.9704	1

从表中可以看出有效就业投资 G' 与就业人数的相关系数0.9578大于总投资 G 与就业人数的相关系数0.9295，进一步证明我们的想法。

5.2.3 仿真模型改进实现

基于以上的分析，我们考虑将有效投资额代替总投资额，改进模型建立考虑有效投资额的模型，将实际的投资总额代换为有效投资额，建立新的多变量VAR模型。然后类似模型一中的仿真过程对有效投资模型进行随机模拟，可以得到2009、2010年的取值概率直方图见图20、21。

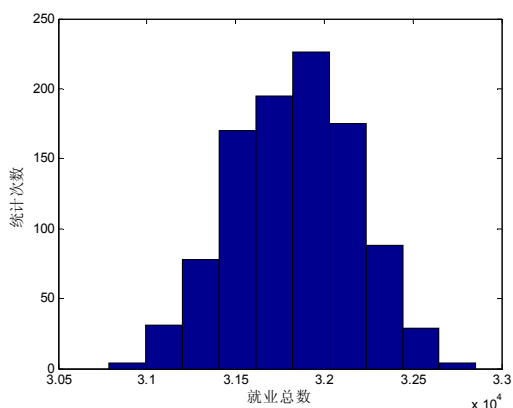


图 20 2009 年城镇就业总人数分布函数图

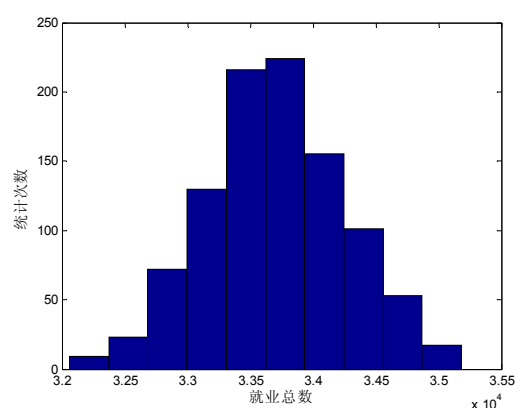


图 21 2010 年城镇就业总人数分布函数图

多次随机模拟得到的图像基本如上图所示分布，显然，都服从正态分布。对以上数

据做正态分布检验，本文用偏度-峰度检验法得到以上两年的数据检验结果见表 17。

表 17 指标预测区间

检验结果	峰度值	偏度值	偏态	均值	方差
2009 年	2.57	-0.039	左偏态	3.18×10^4	0.034×10^4
2010 年	2.79	0.0188	右偏态	3.369×10^4	0.055×10^4

由此可以得知，2009 年的城镇就业总人数在置信度为 95% 的置信区间为 [31000,32750](单位: 万人), 2010 年的总人数在置信度为 95% 的置信区间为 [32500,35000] (单位: 万人)。显然有上述结果可知，使用有效支配投资额建立的模型中经济危机给城镇就业总人数带来的影响不大，且国家投资的 4 万亿投资带来了就业总数的增加。

但是，实际情况下，在一段时间内我国的就业率有大幅度的下降。而以上结果现实就业人数的增加，对于这种现象可以解释为国家宏观调控是整年的就业人数增加。为了更加详细的仿真 2009 年到 2010 年上半年的就业人数变化，本文对模型进一步改进。研究以月份为步长进行仿真。

5.3 考虑经济危机和投资变动的仿真方案三

5.3.1 逐月仿真的微观模型（模型三）的分析

回顾历史，虽然期间爆发过社会动荡和自然灾害，但是从总体上看，我国城镇就业人数总的趋势是稳定上升的：

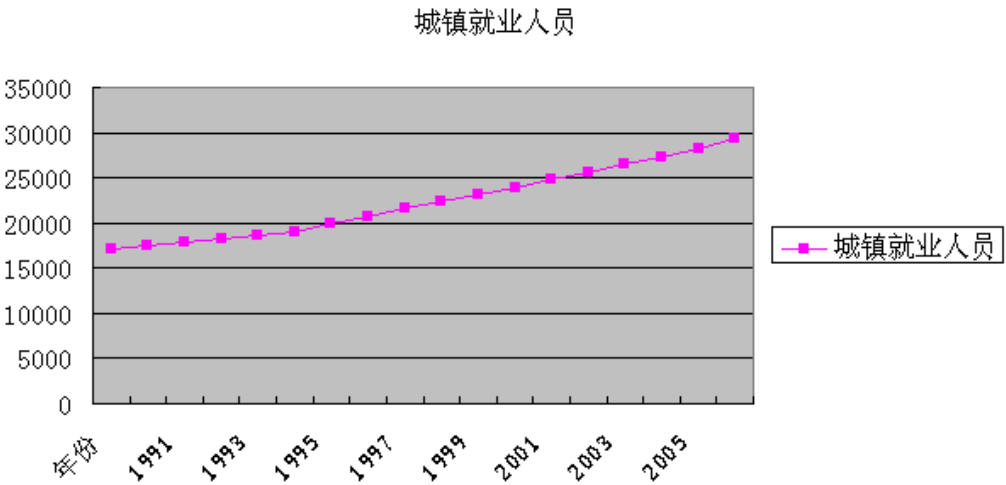


图 22 我国城镇就业人数稳定上升

如果以年作为分析单位，很难看出社会动荡和自然灾害对宏观经济社会的影响，所以本题在此处以月作为分析单位，逐月分析 2008 年至 2010 年 6 月期间我国城镇就业人数和各因素变动情况，其中金融危机的爆发和四万亿投资将是我们研究的重点。

而由于金融危机和四万亿投资存在的“滞后效应”，即金融危机爆发后和投资增加后，不会立刻会我国就业人数产生影响，而一个逐渐的过程。

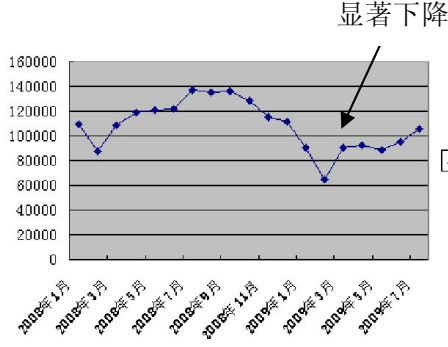


图 23 出口额显著下降

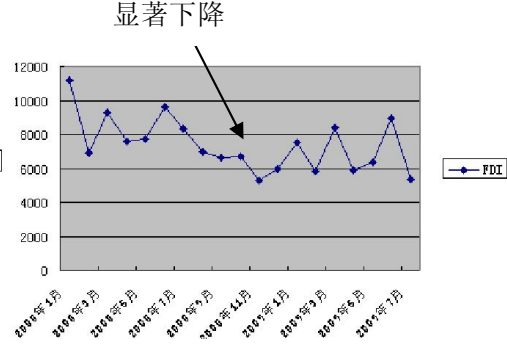


图 24 FDI 显著下降

为了模拟这个过程，我们引入两个影响函数：

P ：金融危机影响函数

Z ：投资影响函数。

函数具体为形式为

$$P = \begin{cases} -\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & t \in [t_0, t_n] \\ 0 & t \notin [t_0, t_n] \end{cases}$$

$$Z = \begin{cases} \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(x-\alpha)^2}{2\beta^2}} & t \in [t_a, t_b] \\ 0 & t \notin [t_a, t_b] \end{cases}$$

关于函数的两点说明：

1. 金融危机影响函数 P 作用时间在 t_0 至 t_n ，投资显著增加的作用时间是在 t_a 至 t_b
2. 在 t_0 至 t_n 时间段内， P 函数作用效果是通过与实际情况拟合出的负高斯分布函数。 t_a 至 t_b 时间段内， Z 函数是根据国家政策调整和相关新闻报道而拟合出的高斯分布函数。

假设在多变量自回归模型，第 k 个变量表示月出口额，第 p 个变量表示月投资金额，则方程可以表示为：

$$\ln x_k(t+1) = m_{k1} \ln x_1(t) + \dots + m_{kn} \ln x_n(t) + P(t) + \varepsilon$$

$$\ln x_p(t+1) = m_{p1} \ln x_1(t) + \dots + m_{pn} \ln x_n(t) + Z(t) + \varepsilon$$

将函数带入以前的模型，以2006年1月为起点，2010年6月为终点，作出不考虑金融危机和四万亿投资的正常预测曲线和考虑金融危机和四万亿投资的仿真曲线，仿真见图25。

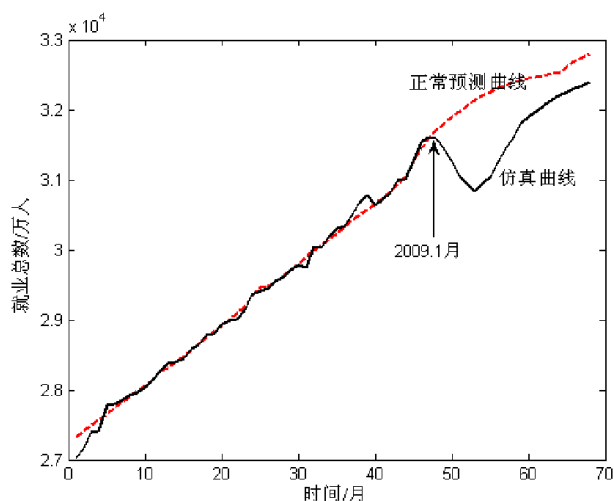


图25 2009年至2010年上半年逐月仿真图

从函数很容易看出，此函数其实是把对我国出口和投资的影响看成是两个部分，并考虑他们不同时期的变化情况，以此来进行对就业人数的模拟。通过模型的求解可以得出：2008年8月份以前金融危机影响函数 P 是不起作用的，8月后，金融危机影响函数 P 开始起作用但不明显，到了2009年1月，由于函数 P 作用，就业人数才开始显著下降，同时，国家进行宏观经济政策的调整，加大投资投资影响函数 Z 开始起作用。

表18 各月城镇就业人数仿真（单位：万人）

月份	就业人数	月份	就业人数
2009年1月	30507	2009年10月	31589
2009年2月	30564	2009年11月	31703
2009年3月	30654	2009年12月	31803
2009年4月	30737	2010年1月	31896
2009年5月	30831	2010年2月	31982
2009年6月	30945	2010年3月	32060
2009年7月	31049	2010年4月	32132
2009年8月	31252	2010年5月	32195
2009年9月	31444	2010年6月	32252

6 问题五解答

——提高城镇就业人口的建议

6.1 投资的分配

6.1.1 投资分配的模型分析

国家增加四万亿投资目的是增加就业，稳定社会生产，促进社会稳定和谐发展，所以投资额的具体分配应该从投资的有效性和公平性两方面综合考虑。投资的效率性即考虑在各个行业的投资是否能最大限度的促进就业，合理性即考虑投资在分配到各个行业时是否考虑到该行业的从业人数和总产值等其他因素。本题通过对各行业情况综合考虑，结合近年的统计数据，建立多目标规划模型对四万亿进行分配。

6.1.1.1 目标函数

1) 公平系数指标：

$$\lambda = \frac{1}{n} \sum_{i=1}^8 \left(\frac{x_i}{s_i p_i} - \frac{\bar{x}}{\bar{s} \bar{p}} \right)^2$$

s_i 表示第 i 个行业的总产值， p_i 表示第 i 个行业的从业人数， x_i 为第 i 个行业的投资额。我们认为各行业的投资额 x_i 应该根据各行业的总产值 s_i ，各行业的从业人数 p_i 客观分配，且各行业的总产值 s_i 越大，各行业的从业人数 p_i 越多，则投资金额越多。

所以对于每个行业 $\frac{x_i}{s_i p_i}$ 应该是尽量相等的，所以我们设定公平系数 $\lambda = \frac{1}{n} \sum_{i=1}^8 \left(\frac{x_i}{s_i p_i} - \frac{\bar{x}}{\bar{s} \bar{p}} \right)^2$ 越小越公平。

则目标函数为：

$$\text{Min} \quad \lambda = \frac{1}{n} \sum_{i=1}^8 \left(\frac{x_i}{s_i p_i} - \frac{\bar{x}}{\bar{s} \bar{p}} \right)^2$$

2) 生产效率指标

三大产业吸收就业能力是不同的，且每个产业中不同行业的吸收就业的能力也是不同的。本文用就业弹性来表示不同产业不同行业的吸收就业能力的不同。参照索洛的新古典经济增长模型，假定存在如下形式的生产函数：

$$Y_t = A_t f(L_t, K_t)$$

式中 Y_t, L_t, K_t 依次代表 t 时期的总产量（或国内生产总值）、投入的劳动量和投入的资本量， A_t 代表 t 时期的技术进步水平，作出如下推导，可以得到：

$$\frac{dY_t}{dt} = \frac{dA_t}{dt} f(L_t, K_t) + A_t \frac{\partial f}{\partial L_t} \cdot \frac{dL_t}{dt} + A_t \frac{\partial f}{\partial K_t} \cdot \frac{dK_t}{dt}$$

在上式两端处以 Y_t ，并且定义参数 $\alpha = \partial Y_t / \partial L_t \cdot L_t / Y_t$ 为劳动的产出弹性，参数 $\beta = \partial Y_t / \partial K_t \cdot K_t / Y_t$ 为资本的产出弹性，则有：

$$\frac{dY_t/dt}{Y_t} = \frac{dA_t/dt}{A_t} + \alpha \frac{dL_t/dt}{L_t} + \beta \frac{dK_t/dt}{K_t}$$

在实际应用时，由于原始资料中 Q 、 L 、 K 均是离散的数据，所以，在时间间隔 Δt 较小时，可以用差分方程近似代替：

$$\frac{\Delta Y_t / \Delta t}{Y_t} = \frac{\Delta A_t / \Delta t}{A_t} + \alpha \frac{\Delta L_t / \Delta t}{\Delta L_t} + \beta \frac{\Delta K_t / \Delta t}{\Delta K_t}$$

当 Δt 趋近单位时间时, 令 $GY = \Delta Y / Y$, $\lambda = \Delta A / A$, $GL = \Delta L / L$, $GK = \Delta K / K$, 则 α 为投资的就业弹性:

$$\alpha = \frac{\Delta Y / Y}{\Delta X / X}$$

该式描述了投资增长与各因素的关系。

测算投资增长的就业弹性意义在于利用这一结果预测四万亿投资对就业的影响, 进而为制定 4 万亿投资去向提供了依据。下表利用公式计算出各行业的投资增长。

表19 各产业各行业就业弹性

第一产业	农业	-0.01296
第二产业	采 矿 业	-0.09442
	制 造 业	-0.00061
	电力、燃气	0.016708
	建 筑 业	0.19339
第三产业	交通运输、仓储、邮政	0.032468
	批发和零售业,住宿和餐饮业	0.029096
	金融业	0.107777
	房地产业	0.164449
	卫生、社会、福利、保障	0.025267
	文化、体育、娱乐	-0.00368
	公共管理和社会组织	-0.04037
	科学研究、技术服务、地质勘探	-0.10141

从而定义效率系数:

$$\eta = \sum_{i=1}^8 \theta_i x_i$$

θ_i 为各行业的投资增长的就业弹性, x_i 为各行业投资金额。我们认为同量的资金, 投入投资增长的就业弹性高的行业比投入投资增长的就业弹性低的行业, 将会吸引更多的人就业。所以补贴应该尽量多的投入在就业弹性高的省区, 效率系数: $\eta = \sum_{i=1}^8 \theta_i x_i$ 越大越好。

则目标函数为: $\text{Max} \quad \sum_{i=1}^8 \theta_i x_i$

6.1.1.2 约束条件

1) 国家四万亿投资金额的约束:

$$\sum_{i=1}^{13} x_i = 40000$$

对 13 个行业的投资总额为 4 万亿元，

2) 各个省区棉花补贴最小数额的约束：

$$x_i \geq 100$$

为考虑补贴分配的公平性与效率性，应保证 13 个行业的投资额不能太大也不能太小，应在[100,15000]亿元范围内变动。

3) 各个省区棉花补贴最大数额的约束

$$x_i \leq 15000$$

为考虑补贴分配的公平性与效率性，应保证 13 个行业的投资额不能太大也不能太小，应在[100,15000]亿元范围内变动。

4) 社会和谐的约束

$$x_{10} \geq 500$$

为了保证社会和谐和稳定发展，卫生、社会保障和社会福利业至少投资 500 亿。

6.1.2 投资分配的模型的建立

目标函数的构成：就业量

$$\text{用 } \lambda \text{ 表示公平系数，则 } \lambda = \frac{1}{n} \sum_{i=1}^8 \left(\frac{x_i}{s_i p_i} - \frac{\bar{x}}{s \bar{p}} \right)^2;$$

$$\text{用 } \eta \text{ 表示效率系数，则 } \eta = \sum_{i=1}^8 \theta_i x_i;$$

公平度越小，表示投资在各行业中分配较公平，效率度越大代表投资分配后能产生最大的。在兼顾公平与效率的基础上，我们可以建立投资分配模型的目标函数为：

$$\text{Max} = \alpha \eta - \beta \lambda$$

其中， α, β 分别为效率偏好系数与公平偏好系数。

得到投资分配的规划模型为：

$$\begin{aligned} \text{Max} &= \alpha \eta - \beta \lambda \\ \text{s.t.} &\begin{cases} \sum_{i=1}^8 x_i \leq 40000 \\ x_i \geq 100 \\ x_i \leq 15000 \\ x_{10} \geq 500 \end{cases} \end{aligned}$$

6.1.3 模型求解与结果分析

我们通过 Lingo 软件求解，并结合各行业的具体情况，得到四万亿投资分配的结果如见表 20。

表20 各产业各行业投资分配表（单位：亿元）

产业	行业	偏好公平	偏好效率
第一产业	农业	6892	100
第二产业	采 矿 业	136	100
	制 造 业	4660	100
	电力、燃气	451	443
	建 筑 业	10234	15000
第三产业	交通运输、仓储、邮政	3156	6098
	批发和零售业,住宿和餐饮业	11972	7077
	金融业	538	2413
	房地产业	730	7570
	卫生、社会、福利、保障	818	799
	文化、体育、娱乐	213	100
	公共管理和社会组织	100	100
	科学研究、技术服务、地质勘探	100	100

偏好公平的系数 $\alpha=0.3, \beta=0.7$ ；偏好效率的系数 $\alpha=0.7, \beta=0.3$

计算结果表明，投资大部分集中于第二产业和第三产业，其中建筑业，房地产等高就业弹性，就业人员密集的行业应该是投资的重点的行业，国家除投资以外，还应该大力扶持相关产业的发展。在偏好公平情况下，应该加大在农业的投资，但随着城镇化和技术进步，依靠农业解决就业问题的思路是行不通的，从宏观角度上，大力发展第二、第三产业是国家将来的投资的趋势。

投资分配决策者可以根据中国就业的实际情况和社会发展水平确定公平与效率系数，作出灵活有效的决策。

6.2 就业人数与投资、出口、和消费的分析

6.2.1 四万亿投资对就业人数的影响

在第四问中，通过对2009年至2010年上半年逐月仿真，发现由于金融危机发生和投资的滞后性，会导致大量人员失业。本问将以投资，出口和GDP作为研究对象，定量分析就业人数与其关系。

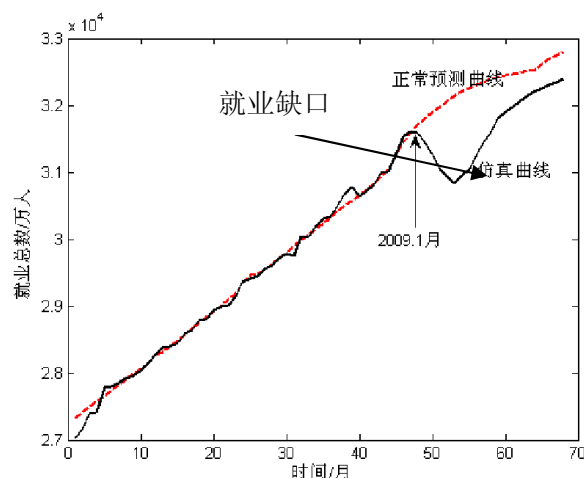


图26 2009年至2010年上半年逐月仿真图

在第四问中，根据所建立的数学模型，仿真出假设没有发生金融危机，2009年我国正常城镇就业人数为31343万人，但受到金融危机的影响，造成大量失业。这里假设四

万亿投资全部在 2009 年和 2010 年这两年内投资完毕且其他宏观经济条件无重大波动。则通过模型仿真得到结果如下表：

表 21 2009 和 2010 年投资额与就业缺口关系表（单位：万人）

	投资情况	就业缺口（区间）	填补缺口所需投资
2009 年	不投资	[650,850]	2.5 万亿
	投资 2 万亿	[100,160]	
2010 年	不投资	[1600,1800]	4 万亿
	投资 2 万亿	[500,700]	

通过表可以得出一下结论：

- 1.如果不进行增加四万亿投资，则 2009 年会有 650-850 万人的就业缺口，而 2010 年，就业缺口将会达到 1600-1800 万人。
- 2.如果每年投资 2 万亿元，则 2009 年会有 100-160 万人的就业缺口，而 2010 年，就业缺口将会达到 500-700 万人。
- 3.假设完全弥补就业缺口，则 2009 年需要 2.5 万亿投资，2010 年需要 4 万亿投资，则四万亿无法弥补就业缺口，总投资到达 6.5 万亿才可以弥补就业缺口。

6.2.2 消费和出口对就业人数的影响

接下来我们在 2009 年和 2010 年每年各增加投资两万亿的假设下，进一步深入研究消费和出口和就业人数的影响。

表 21 2009-2010 每年增加投资两万亿就业分析表

年份	就业缺口	弥补缺口所需增加消费额	弥补缺口所需增加出口额
2009	100-160 万人	166-210（人/元）	900-1000（亿人民币）
2010	500-700 万人	700-800（人/元）	4000-4500（亿人民币）

通过表可以得出一下结论：

- 1.在每年增加投资两万亿的前提下，2009 年人均只要增加 166-210 元消费，即可填补就业缺口，而 2010 年人均需要增加 700-800 元消费，可填补就业缺口。
- 2.在每年增加投资两万亿的前提下，2009 年出口金额增加 900-1000 亿人民币，可填补就业缺口，而 2010 年出口金额增加 4000-4500 亿人民币，可填补就业缺口。

6.2.3 咨询与建议

1. 扩大内需

在国外市场需求不足，金融危机的大背景下，我国出口连续 7 个月负增长，所以扩大出口额的可能性很小，这时国内投资和内需增长显得尤为重要。在每年增加投资两万亿的前提下，2009 年人均只要增加 166-210 元消费，2010 年人均需要增加 700-800 元消费，可填补就业缺口。在国外需求不足的情况下，扩大内需是解决解决就业的有效的途径，国家可以通过增加工资或货币流通量来刺激消费，扩大内需。

2. 增加投资

假设其他宏观经济条件不变的情况下，经计算得出在 2009 年至 2010 年之间共投资 6.5 亿可以填补就业缺口，这说明四万亿投资并不能完全解决失业问题，但是可以大大减少失业人数，对社会的稳定和和谐起到了重大作用，四万亿投资是及时有效的。如果想完全解决就业缺口，需要增加投资。

3. 加大对第二，第三产业的投资

通过多目标规划模型求解结果表明，无论是偏重公平还是偏重效率，第二产业和第三产业都是投资的重点产业。在我国，第二，第三产业提供了大量工作岗位，且第三产

业还有很大的发展空间。所以国家在分配投资资金的时候应该向第二产业，第三产业倾斜。

7 模型评价和推广

7.1 模型的优点

1. 斯皮尔曼 (Spearman) 相关系数对数据不要求服从正态分布，且可以反应变量间趋同关系。
2. 多变量自回归对数线性模型准确度高，稳定性好，能较好地克服多变量的共线性问题。
3. 路径分析可以清晰明了地反映多变量间层次关系和相关关系。
4. 多目标规划可以综合考虑，根据决策者要求和实际情况变化作出相应的决策。

7.2 模型的缺点

1. 前提假设过多，可能有些与实际情况不符。
2. 斯皮尔曼 (Spearman) 相关系数准确度不高。
3. 路径分析的结果只能反映数值上的关系，不能反映逻辑上的关系。

7.3 模型的推广

1. 多变量自回归对数线性模型可以推广到变量较多且相互之间存在相互作用的情况，例如天气，股票预测，工程等领域中，准确度高且稳定性好。
2. 该模型可以推广到解决多个变量有关联，而且有滞后性的问题，并应用于实际问题中的预测问题。

参考文献

- [1] 沈全水.失业的出路:瑞典就业政策及其对中国的启示[M].北京:中国发展出版社,2000.
- [2] 裘元伦、罗红波.中国与欧洲联盟就业政策比较[M]. 北京:中国经济出版社,1998.
- [3] 熊建国.安徽省经济增长与劳动力就业的实证分析[J].区域经济, 2006
- [4] 于俊平.计量经济学[M].北京: 对外经济贸易大学出版社, 2000
- [5] 胡鞍钢. 中国就业状况分析[J]. 管理世界,1997(5)
- [6] 范建勇. 改革以来就业结构变动及其对经济增长的贡献[J].宏观经济研究, 2001(9)
- [7] 梁庆文. 我国人口就业增长与经济发展状况的关系分析[J].数理统计与管理, 2000(5)
- [8] [美]萨缪尔森,诺德豪斯著,萧琛译. 经济学(第十六版) [M]. 北京:华夏出版社,1999
- [9] 胡乃武,金碚. 国外经济增长理论比较研究[M]. 北京:中国人民大学出版社,1990
- [10]成学真 郑贺娟.经济增长对就业的影响---理论研究综述[J].经济研究, 2005

- [11]张莉,张德然.关于四川就业问题的预测模型的比较研究[J].商丘师范学院学报,2008
- [12]王婧.基于因果关系理论与BP神经网络整合模型的就业研究,西安建筑科技大学博士论文,2008
- [13]赵慧,刘德鑫,张家来.经济增长与扩大劳动力就业的相关问题分析[J].当代财经,2007(6)
- [14]冯政.中国就业形势和积极的就业政策[EB/OL].中国劳动力市场信息网监测中心,2004-08-17.
- [15]李文花.我国经济增长与就业增长的关系[J].东疆学刊,2005,(1).
- [16]孙祖芳.关业就业与失业问题的理论政策研究[J].社会科学,2004,(3).
- [17]丁乐臣.基于CGE模型的我国劳动就业形势分析,博士论文,2009
- [18]蔡昉.就业弹性、自然失业和宏观经济政策[J].经济研究,2004,(9).
- [19]李俊峰,邹红美.就业与经济增长的相关性分析[J].统计观察,2005,(4).
- [20]陈阳.辽宁省就业和经济增长关系的时变参数模型[J].统计与信息论坛,2007,(9).
- [21]张朝黎.影响我国劳动力就业状况的诸因素分析[J].现代财经,2000,(9).
- [22]谭砚文等.中日美服务业劳动生产率对经济增长促进作用的比较分析[J].数量经济技术经济研究,2007,(12).
- [23]刘萍.当前我国就业状况及对策分析[J].福建财会管理干部学院学报,2004年第1期
- [24]胡鞍钢.就业与发展:中国失业问题与就业战略[M],沈阳:辽宁人民出版社,1998.
- [25]《技术进步对我国就业总量及结构的影响》 毕先萍^{1,2},赵坚毅³
- [26]《FDI对中国就业影响的动态分析》 钟辉 [J]世界经济研究所,2005

附录

1、模型建立程序

```
function ti_2()
clc
A=load('D:\ti2.txt');
temp_A=A;
[m,n]=size(A);
corr(A)
det(corr(A))
%%%%%%%%%确定六个微分方程的系数
AA=A(1:end-1,:);
%%%%%%%%%建立方程一
b=A(2:end,1); A_ni=inv(AA'*AA); y1=A_ni*AA'*b;
%%%%%%%%%建立方程二
b=A(2:end,2); A_ni=inv(AA'*AA); x1=A_ni*AA'*b;
%%%%%%%%%建立方程三
b=A(2:end,3); A_ni=inv(AA'*AA); x2=A_ni*AA'*b;
%%%%%%%%%建立方程四
b=A(2:end,4); A_ni=inv(AA'*AA); x3=A_ni*AA'*b;
%%%%%%%%%建立方程五
b=A(2:end,5); A_ni=inv(AA'*AA); x4=A_ni*AA'*b;
%%%%%%%%%建立方程六
b=A(2:end,6); A_ni=inv(AA'*AA); x5=A_ni*AA'*b;
%%%%%%%%%建立方程七
b=A(2:end,7); A_ni=inv(AA'*AA); x6=A_ni*AA'*b;
%%%%%%%%%55建立方程8
b=A(2:end,8); A_ni=inv(AA'*AA); x7=A_ni*AA'*b;
%%%%%%%%%55建立方程8
b=A(2:end,9); A_ni=inv(AA'*AA); x8=A_ni*AA'*b;
r_x=[y1,x1,x2,x3,x4,x5,x6,x7,x8];
%%可得到差分方程组,用上一年数据预测下一年的数据
for i=1:m
    for j=1:n
        nihe(i,j)=temp_A(i,:)*r_x(:,j);
    end
end
temp_A;
nihe=[temp_A(1,:);nihe]
figure(1);
subplot(3,3,1);plot([1990:2006],A(:,1),'');hold on;plot([1990:2006+1],nihe(:,1),'r');legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('就业人数/万人');
subplot(3,3,2);plot([1990:2006],A(:,2),'');hold on;plot([1990:2006+1],nihe(:,2),'r');%legend('原始数据','拟合曲线')
```



```

线')
xlabel('时间/年');ylabel('GDP/亿元');
subplot(3,3,3);plot([1990:2006],A(:,3),'');hold on;plot([1990:2006+1],nihe(:,3),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('城镇总投资/亿元');
subplot(3,3,4);plot([1990:2006],A(:,4),'');hold on;plot([1990:2006+1],nihe(:,4),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('教育投资/亿元');
subplot(3,3,5);plot([1990:2006],A(:,5),'');hold on;plot([1990:2006+1],nihe(:,5),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('出口贸易额/亿元');
subplot(3,3,6);plot([1990:2006],A(:,6),'');hold on;plot([1990:2006+1],nihe(:,6),'r');
xlabel('时间/年');ylabel('城镇平均工资/%');
subplot(3,3,7);plot([1990:2006],A(:,7),'');hold on;plot([1990:2006+1],nihe(:,7),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('市场化程度/元/每人');
subplot(3,3,8);plot([1990:2006],A(:,8),'');hold on;plot([1990:2006+1],nihe(:,8),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('人口总数/万人');
subplot(3,3,9);plot([1990:2006],A(:,9),'');hold on;plot([1990:2006+1],nihe(:,9),'r');%legend('原始数据','拟合曲线')
xlabel('时间/年');ylabel('利率/%');

```

2、仿真程序

```

function fangzheng_2()%第三问的仿真模型的改进
%主要考虑经济危机带来的突变。本文在本问考虑 投资、进出口额 的变化。
%%%%%%%%%%%%%%对于仿真没有考虑因素，通过实际的历史数据，
仿真过程中通过预测取得
%%%%%%%%%%其中不变因素有1、GDP总额 2、教育经费 3、市场化程度 4、人口总量 5、利率 6、城
镇人口比例
%%时间预测2009年以上1-6的变量系数的取值区间
clc;clear;
A=load('D:\ti3.1.txt');
[m,n]=size(A);lilv=0.27;
GDP=[141552.4093    256675.0462;
     126957.9853    266050.14];
edu=[1696.3482    2771.5719;1563.1351    2869.3918;];
renkou=[129814.0    134299.9;128826.2    135144.7;];
shichang=[5798.1496    7661.7572;5381.8078    8000.9788;];
chengzhen=[30.6572    35.7360;30.0324    36.2105;];
%%%%%%%%%%%%%%5两个特殊的指标
touzi=[26003.3099    35757.0204;
       23810.3053+40000*0.154    37501.5280+40000*0.154];%有效投资额

```

```

chukou=[14291.7,14291.7;
        11785.8,11148];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
die=1000;
for i=1:die %迭代随机求的2009年后就业总人数
    y=30000;
    x1=GDP(1,1)+rand*(GDP(1,2)-GDP(1,1));%GDP
    xx1=GDP(2,1)+rand*(GDP(2,2)-GDP(2,1));%GDP
    x2=touzi(1,1)+rand*(touzi(1,2)-touzi(1,1));%总投资额
    xx2=touzi(2,1)+rand*(touzi(2,2)-touzi(2,1));%总投资额
    x3=edu(1,1)+rand*(edu(1,2)-edu(1,1));%教育经费
    xx3=edu(2,1)+rand*(edu(2,2)-edu(2,1));%教育经费
    x4=chukou(1,1)+rand*(chukou(1,2)-chukou(1,1));%出口贸易额
    xx4=chukou(2,1)+rand*(chukou(2,2)-chukou(2,1));%出口贸易额
    x5=shichang(1,1)+rand*(shichang(1,2)-shichang(1,1));%市场程度
    xx5=shichang(2,1)+rand*(shichang(2,2)-shichang(2,1));%市场程度
    x6=renkou(1,1)+rand*(renkou(1,2)-renkou(1,1));%人口总量
    xx6=renkou(2,1)+rand*(renkou(2,2)-renkou(2,1));%人口总量
    x7=lilv;xx7=lilv;%利率
    x8=chengzhen(1,1)+rand*(chengzhen(1,2)-chengzhen(1,1));%城镇化水平
    xx8=chengzhen(2,1)+rand*(chengzhen(2,2)-chengzhen(2,1));%城镇化水平
    jiuye(1,i)=0.85502*y+0.0028768*x1+0.085243*x2-0.45296*x3+0.065377*x4+0.10753*x5+0.046523*x6-
    106.97*x7-114.29*x8;
    yy=jiuye(1,i);
    jiuye(2,i)=0.85502*yy+0.0028768*xx1+0.085243*xx2-0.45296*xx3+0.065377*xx4+0.10753*xx5+0.0465
    23*xx6-106.97*xx7-114.29*xx8;
end
[mean(jiuye'),std(jiuye')]
[R1,R2]=hist(jiuye) ;
figure(1);
hist(jiuye(1,:));
xlabel('就业总数');ylabel('统计次数');
figure(2);
hist(jiuye(2,:));
xlabel('就业总数');ylabel('统计次数');
jianyan=[skewness(jiuye(1,:)),kurtosis(jiuye(1,:));
skewness(jiuye(2,:)),kurtosis(jiuye(2,:));]
% w=[2.93:0.025:3.175];
% R1,w
% hold on;
% plot(w,R1) ,
% fplot('1/(sqrt())*exp()^2');
% hold on
% fplot();

```

3、路径分析程序

```

data path;
input y x1-x16@@;
cards;
1    0.94351 0.92036 0.9979  0.98808 0.98808 -0.51708 -0.5178  0.95537 0.93008 0.97412
    0.93443 0.97341 0.99936 0.71084 -0.81818 0.98459
0.94351 1    0.96294 0.93222 0.94517 0.94517 -0.40136 -0.40225 0.99823 0.99876 0.95549
    0.84582 0.86318 0.93854 0.52279 -0.70576 0.9683
0.92036 0.96294 1    0.9023  0.9431  0.9431  -0.35914 -0.35999 0.97303 0.9541  0.97464
    0.87305 0.83633 0.90881 0.50369 -0.65993 0.97234
0.9979  0.93222 0.9023  1    0.98726 0.98726 -0.51494 -0.51562 0.94447 0.91797 0.96741
    0.94159 0.98238 0.99819 0.74289 -0.8284  0.97632
0.98808 0.94517 0.9431  0.98726 1    1    -0.43199 -0.43274 0.96042 0.92927 0.99065
    0.96612 0.96738 0.98382 0.72934 -0.78827 0.99042
0.98808 0.94517 0.9431  0.98726 1    1    -0.43199 -0.43274 0.96042 0.92927 0.99065
    0.96612 0.96738 0.98382 0.72934 -0.78827 0.99042
-0.51708 -0.40136 -0.35914 -0.51494 -0.43199 -0.43199 1    1    -0.40536 -0.39479 -0.43593
    -0.36741 -0.47289 -0.52239 -0.48014 0.76597 -0.46345
-0.5178  -0.40225 -0.35999 -0.51562 -0.43274 -0.43274 1    1    -0.40625 -0.39569 -0.43667
    -0.36804 -0.47351 -0.5231  -0.48038 0.76637 -0.46423
0.95537 0.99823 0.97303 0.94447 0.96042 0.96042 -0.40536 -0.40625 1    0.99413 0.97049
    0.87205 0.88175 0.94958 0.55138 -0.71691 0.98007
0.93008 0.99876 0.9541  0.91797 0.92927 0.92927 -0.39479 -0.39569 0.99413 1    0.94071
    0.81998 0.84272 0.92568 0.49256 -0.69226 0.95635
0.97412 0.95549 0.97464 0.96741 0.99065 0.99065 -0.43593 -0.43667 0.97049 0.94071 1
    0.94949 0.93156 0.96655 0.66822 -0.7617  0.99521
0.93443 0.84582 0.87305 0.94159 0.96612 0.96612 -0.36741 -0.36804 0.87205 0.81998
    0.94949 1    0.96267 0.92746 0.81446 -0.7379  0.93069
0.97341 0.86318 0.83633 0.98238 0.96738 0.96738 -0.47289 -0.47351 0.88175 0.84272
    0.93156 0.96267 1    0.97478 0.81842 -0.80423 0.93388
0.99936 0.93854 0.90881 0.99819 0.98382 0.98382 -0.52239 -0.5231  0.94958 0.92568
    0.96655 0.92746 0.97478 1    0.71464 -0.82368 0.97923
0.71084 0.52279 0.50369 0.74289 0.72934 0.72934 -0.48014 -0.48038 0.55138 0.49256
    0.66822 0.81446 0.81842 0.71464 1    -0.79612 0.65117
-0.81818 -0.70576 -0.65993 -0.8284  -0.78827 -0.78827 0.76597 0.76637 -0.71691 -0.69226 -0.7617
    -0.7379  -0.80423 -0.82368 -0.79612 1    -0.77858
0.98459 0.9683  0.97234 0.97632 0.99042 0.99042 -0.46345 -0.46423 0.98007 0.95635
    0.99521 0.93069 0.93388 0.97923 0.65117 -0.77858 1
;
proc print;
run;
proc reg; model y=x1-x16;
run;

```